(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 July 2002 (04.07.2002)

PCT

(10) International Publication Number
WO 02/051857 A1

(51) International Patent Classification[7]: C07H 21/02, C12Q 1/68, C12P 21/00, 19/34, A01N 43/04, G06F 17/00

(21) International Application Number: PCT/US00/34758

(22) International Filing Date:
21 December 2000 (21.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 0511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHEW, Anne** [US/US]; 1477 Beacon Street #64, Brookline, MA 02446 (US). **DENTON, R., Rex** [US/US]; 129 Hunter's Trail, Madison, CT 06443 (US). **LANZ, Elizabeth, M.** [US/US]; 111 Park Street, No. 5U, New Haven, CT 06511 (US). **NANDABALAN, Krishnan** [IN/US]; 228 Village Pond Road, Guilford, CT 06437 (US). **STEPHENS, J., Claiborne** [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US).

(74) Agents: HENDERSON, Melodie, W. et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/051857 A1

(54) Title: DRUG TARGET ISOGENES: POLYMORPHISMS IN THE NEUROPEPTIDE Y GENE

(57) Abstract: Polynucleotides comprising one or more of 8 novel single nucleotide polymorphisms in the human Neuropeptide Y (NPY) gene are described. Compositions and methods for detecting one or more of these polymorphisms are also disclosed. In addition, various genotypes and haplotypes for the NPY gene that exist in the population are described.

DRUG TARGET ISOGENES: POLYMORPHISMS IN THE NEUROPEPTIDE Y GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/171,414 filed on December 21, 1999.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically important proteins. In particular, this invention provides genetic variants of the human Neuropeptide Y (NPY) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a drug that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by a compound's activity at non-intended targets.

What this approach fails to consider, however, is that natural variability exists in any and every population with respect to a particular protein. A target protein currently used to screen drugs typically is expressed by a gene cloned from an individual who was arbitrarily selected. However, the nucleotide sequence of a particular gene may vary tremendously among individuals. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a target protein may be manifested as significant variation in expression of or in the structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in treatment of individuals with drugs whose design is based upon a single representative example of the target. For example, it is well-established that some classes of drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. In addition, variable information on the biological function or effects of a particular protein may be due to different scientists unknowingly studying different isoforms of the gene encoding the protein. Thus, information on the type and frequency of genomic variation that exists for pharmaceutically important proteins would be useful.

The organization of single nucleotide variations (polymorphisms) in the primary sequence of a gene into one of the limited number of combinations that exist as units of inheritance is termed a

haplotype. Each haplotype therefore contains significantly more information than individual unorganized polymorphisms. Haplotypes provide an accurate measurement of the genomic variation in the two chromosomes of an individual and can be used to identify and distinguish between isoforms of genes coding for drug targets.

Many diseases are associated with specific variations in gene sequences. However while there are examples in which individual polymorphisms act as genetic markers for a particular phenotype, in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74). Because a haplotype represents the variation across each form of a gene, it is superior to a SNP, or even multiple unordered SNPs in the gene, as a tool to determine whether a particular gene is involved in a particular phenotype. For example, analysis of the association between a particular phenotype and each observed haplotype for a gene suspected to be associated with that particular phenotype permits ranking of each haplotype by its statistical power of prediction for the phenotype. Haplotypes found to be strongly associated with the phenotype can then have that positive association confirmed by alternative methods to minimize false positives. Conversely, if no observed haplotypes for that gene show association with the phenotype of interest, then it may be inferred that variation in the gene has little, if any, involvement with that phenotype. Thus, the ability to determine readily whether a gene is involved in the cause or symptoms of a particular disease, or in the response to a particular drug, will be one useful application for haplotypes in development of diagnostics and drugs.

One possible drug target for the treatment of atherosclerosis, obesity, psychological disorders and alcoholism is the Neuropeptide Y (NPY) gene or its encoded product. NPY is a 36-amino acid vasoactive peptide widely distributed throughout the central and peripheral nervous system (Ito, H. et al., Clin. Nephrol. 51:272-279, 1999; Karvonen, M.K. et al., Nat. Med. 4:1434-1437, 1998; Ann. Med 30:508-510, 1998). NPY has a vasocontrictor effect and enhances the effect of various agonists, including noradrenaline and angiotensin II (Pedrazzini et al., Curr Opin Nephrol Hypertens 1993 2(1):106-13). NPY plays an important role in the contraction of heart muscle by inducing an increase in the levels of cytosolic and nuclear calcium in heart and vascular smooth muscle cells (Jacques et al., Can J. Physiol Pjarmacol 2000 78(2):162-72). NPY is a potent orexigenic agent and is believed to be involved in the regulation of eating behavior (Kokot and Ficek., Miner Electrolyte Metab 1999 25(4-6):303-5). It is also involved in the regulation of neuronal activity both under physiological conditions and during pathological hyperactivity such as seizures. Experimental models and tissue from epilepsy sufferers suggest that NPY-mediated neurotrasmission is altered in seizures (Vezzani et al. 1999 22(1):25-30).

The Neuropeptide Y gene is located on chromosome 7pter-q22 and contains 3 exons that encode a 97 amino acid protein. Reference sequences for the NPY gene (reverse complement of a

portion of GenBank Accession No: AC004485.1; SEQ ID NO:1), coding sequence (GenBank Accession No. NM_000905.1), and protein are shown in Figures 1, 2 and 3, respectively.

Several single nucleotide polymorphisms (SNPs) in the NPY gene have been reported in the NCBI SNP Database which correspond to the following polymorphisms in Figure 1: a cytosine or adenine at a position corresponding to nucleotide 155413 (Ref SNP ID# 16146), a guanine or thymine at a position corresponding to nucleotide 155420 (Ref SNP ID# 16145), a guanine or adenine at a position corresponding to nucleotide 155562 (Ref SNP ID# 16144), a cytosine or thymine at a position corresponding to nucleotide 156019 (Ref SNP ID#16143), a guanine or adenine at a position corresponding to nucleotide 156039 (Ref SNP ID# 16478), an adenine or guanine at a position corresponding to nucleotide 156101 (Ref SNP ID# 16142), a thymine or cytosine at a position corresponding to nucleotide 156310 (Ref SNP ID#16139), a guanine or adenine at a position corresponding to nucleotide 156374 (Ref SNP ID# 5572), an adenine or guanine at a position corresponding to nucleotide 162782 (Ref SNP ID#5576), an adenine or guanine at a position corresponding to nucleotide 162917 (Ref SNP ID# 16475) and a thymine or cytosine at a position corresponding to nucleotide 162948 (Ref SNP ID# 16126).

The polymorphism of thymine or cytosine at nucleotide position 156310 in Fig. 1 results in an amino acid variation of leucine (Leu) or proline (Pro) at amino acid position 7 in Figure 3. An association has been identified between the polymorphism producing the Pro7 variant and higher serum levels of total and LDL cholesterol in obese Finnish and Dutch subjects (Karvonen et al., *supra*). Nishanen et al. (J. Clin. Endocrinol Metab 2000 86(6):2266-9) also reported that subjects having Pro7 had higher serum cholesterol and apolipoprotein B levels, and are at a higher risk for carotid artherosclerosis than individuals with the Leu7 variant. Recent studies on middle-aged men from Eastern Finland have shown that the Leu7Pro variation is associated with alcohol consumption. Individuals with Pro7 showed a higher average alcohol consumption than individuals with Leu7 (Kauhanen et al., Am. J. Med. Genet 2000 93(2):117-21). Because of the potential for polymorphisms in the NPY gene to affect the expression and function of the encoded protein, it would be useful to determine whether additional polymorphisms exist in the NPY gene, as well as how such polymorphisms are combined in different copies of the gene. Such information would be useful for studying the biological function of NPY as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 8 novel polymorphic sites in the NPY gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the reverse complement of the indicated GenBank Accession Number: 154224 (PS1), 154841 (PS2), 155063 (PS3), 155119 (PS4), 155412 (PS5), 155954 (PS9), 156117 (PS13) and 156259 (PS14) in AC004485.1. The polymorphisms at these sites are thymine or guanine at PS1, thymine or cytosine

at PS2, guanine or thymine at PS3, cytosine or adenine at PS4, guanine or thymine at PS5, thymine or cytosine at PS9, adenine or guanine at PS13 and adenine or cytosine at PS14. In addition, the inventors have determined the identity of the alternative nucleotides present at these sites, as well as at the previously identified sites at nucleotides 155413 (PS6), 155420 (PS7), 155562 (PS8),156019 (PS10), 156039 (PS11), 156101 (PS12), 156310 (PS15), 156374 (PS16), 162782 (PS17), 162917 (PS18) and 162948 (PS19), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. It is believed that NPY-encoding polynucleotides containing one or more of the novel polymorphic sites reported herein will be useful in studying the expression and biological function of NPY, as well as in developing drugs targeting this protein. In addition, information on the combination of polymorphisms in the NPY gene may have diagnostic and forensic applications.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the NPY gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, thymine at PS5, cytosine at PS9, guanine at PS13 and cytosine at PS14. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of adenine at PS6, thymine at PS7, adenine at PS8, thymine at PS10, adenine at PS11, guanine at PS12, cytosine at PS15, adenine at PS16, guanine at PS17, guanine at PS18 and cytosine at PS19. A particularly preferred polymorphic variant is a naturally-occurring isoform (also referred to herein as an "isogene") of the NPY gene. A NPY isogene of the invention comprises thymine or guanine at PS1, thymine or cytosine at PS2, guanine or thymine at PS3, cytosine or adenine at PS4, guanine or thymine at PS5, cytosine or adenine at PS6, guanine or thymine at PS7, guanine or adenine at PS8, thymine or cytosine at PS9, cytosine or thymine at PS10, guanine or adenine at PS11, adenine or guanine at PS12, adenine or guanine at PS13, adenine or cytosine at PS14, thymine or cytosine at PS15, guanine or adenine at PS16, adenine or guanine at PS17, adenine or guanine at PS18 and thymine or cytosine at PS19. The invention also provides a collection of NPY isogenes, referred to herein as a NPY genome anthology.

A NPY isogene may be defined by the combination and order of these polymorphisms in the isogene, which is referred to herein as a NPY haplotype. Thus, the invention also provides data on the number of different NPY haplotypes found in the above four population groups. This haplotype data is useful in methods for deriving a NPY haplotype from an individual's genotype for the NPY gene and for determining an association between a NPY haplotype and a particular trait.

Polynucleotides complementary to these NPY genomic DNA variants are also provided by the invention.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well

as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express NPY for protein structure analysis and drug binding studies.

In other embodiments, the invention provides methods, compositions, and kits for haplotyping and/or genotyping the NPY gene in an individual. The methods involve identifying the nucleotide or nucleotide pair present at one or more polymorphic sites selected from PS1-PS5, PS9, PS13 and PS14 in one or both copies of the NPY gene from the individual. The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic site. The methods and compositions for establishing the genotype or haplotype of an individual at the novel polymorphic sites described herein are useful for studying population diversity, anthropological lineage, the significance of diversity and lineage at the phenotypic level, paternity testing, forensic applications, and for identifying associations between the NPY genetic variation and a trait such as level of drug response or susceptibility to disease.

In yet another embodiment, the invention provides a method for identifying an association between a genotype or haplotype and a trait. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for atherosclerosis, obesity, psychological disorders and alcoholism.

The present invention also provides nonhuman transgenic animals comprising one of the NPY polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the NPY isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against NPY protein, and for testing the efficacy of therapeutic agents and compounds for atherosclerosis, obesity, psychological disorders and alcoholism in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the NPY gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the NPY gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing NPY haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the NPY gene that is the reverse complement of part of Genbank Accession Number AC004485.1 (contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated below the sequence by the numbers within the brackets and the polymorphic sites and polymorphisms identified by Applicants in a

reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:89 is equivalent to Figure 1, with the two allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (r= g or a, y= t or c, m= a or c, k= g or t, s= g or c, and w= a or t; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the NPY coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site and polymorphism identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the NPY protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Fig. 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the NPY gene. As described in more detail below, the inventors herein discovered 8 novel polymorphic sites by characterizing the NPY gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals) Asian (20 individuals) Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

Using the NPY genotypes identified in the Index Repository and the methodology described in the Examples below, the inventors herein also determined the haplotypes found on each chromosome for most human members of this repository. The NPY genotypes and haplotypes found in the repository include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for studying population diversity, anthropological lineage, the significance of diversity and lineage at the phenotypic level, paternity testing, forensic applications, and for identifying associations between the NPY genetic variation and a trait such as level of drug response or susceptibility to disease.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of

an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g.,

naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

The inventors herein have discovered 8 novel polymorphic sites in the NPY gene. The polymorphic sites identified by the inventors are referred to as PS1-19 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1-PS5, PS9, PS13 and PS14.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the NPY gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant NPY gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1-PS5, PS9, PS13 and PS14, and may also comprise one or more additional polymorphisms selected from the group consisting of adenine at PS6, thymine at PS7, adenine at PS8, thymine at PS10, adenine at PS11, guanine at PS12, cytosine at PS15, adenine at PS16, guanine at PS17, guanine at PS18 and cytosine at PS19. Similarly, the nucleotide sequence of a variant fragment of the NPY gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence (or other reported NPY sequences) or to portions of the reference sequence (or other reported NPY sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of guanine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, thymine at PS5, cytosine at PS9, guanine at PS13 and cytosine at PS14. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the NPY gene which is defined by any one of haplotypes 1-20 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the NPY gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

NPY isogenes may be isolated using any method that allows separation of the two "copies" of the NPY gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in

Ruaño et al., Proc. Natl. Acad. Sci. 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 17 Nucleic Acids. Res. 8392, 1989; Ruaño et al., 19 Nucleic Acids Res. 6877-6882, 1991; Michalatos-Beloin et al., 24 Nucleic Acids Res. 4841-4843, 1996).

The invention also provides NPY genome anthologies, which are collections of NPY isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A NPY genome anthology may comprise individual NPY isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the NPY isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred NPY genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded NPY protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant NPY sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors,

herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 Science 282:1145-1147). Particularly preferred host cells are mammalian cells.

Polymorphic variants of fragments according to the invention comprise at least one novel polymorphism identified herein and have a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the NPY gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the NPY genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular NPY protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the NPY isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular NPY isogene. Expression of a NPY isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of NPY mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of NPY mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to

increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

Effect(s) of the polymorphisms identified herein on expression of NPY may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant of the NPY gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into NPY protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired NPY isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the NPY isogene is introduced into a cell in such a way that it recombines with the endogenous NPY gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired NPY gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the NPY isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the NPY isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant organisms, i.e., transgenic animals, expressing a variant NPY gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the NPY isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In:

Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human NPY isogene and producing human NPY protein can be used as biological models for studying diseases related to abnormal NPY expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel NPY isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel NPY isogenes; an antisense oligonucleotide directed against one of the novel NPY isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel NPY isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel NPY isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Information on the identity of genotypes and haplotypes for the NPY gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of basic research and clinical applications. Thus, the invention also provides compositions and methods for detecting the novel NPY polymorphisms

identified herein.

The compositions comprise at least one NPY genotyping oligonucleotide. In one embodiment, a NPY genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a NPY polynucleotide, i.e., a NPY isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-NPY polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the NPY gene using the polymorphism information provided herein in conjunction with the known sequence information for the NPY gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where

such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruano et al., 87 Proc. Natl. Acad. Sci. USA 6296-6300, 1990. Typically, an allele-specific oligonucleotide will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotide probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15 mer, the 8th or 9th position in a 16mer, the 10th or 11th position in a 20 mer). A preferred ASO probe for detecting NPY gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TATTCATTCAACAGG (SEQ ID NO:4) and its complement,
TATTCATGCAACAGG (SEQ ID NO:5) and its complement,
CAGTGGGTTGGTAGT (SEQ ID NO:6) and its complement,
CAGTGGGCTGGTAGT (SEQ ID NO:7) and its complement,
CGGGAGGGTTGGGGT (SEQ ID NO:8) and its complement,
CGGGAGGTTTGGGGT (SEQ ID NO:9) and its complement,
CAGCCCCCTCCCCCC (SEQ ID NO:10) and its complement,
CAGCCCCATCCCCCC (SEQ ID NO:11) and its complement,
CTGTGGAGCCCTGGG (SEQ ID NO:12) and its complement,
CTGTGGATCCCTGGG (SEQ ID NO:13) and its complement,
AGAAAAGTGACCCAG (SEQ ID NO:14) and its complement,
AGAAAAGCGACCCAG (SEQ ID NO:15) and its complement,
CGCCCGGAGCCCGCA (SEQ ID NO:16) and its complement,
CGCCCGGGGCCCGCA (SEQ ID NO:17) and its complement,
AATCCCCAAGCCCGT (SEQ ID NO:18) and its complement, and
AATCCCCCAGCCCGT (SEQ ID NO:19) and its complement.

An allele-specific oligonucleotide primer of the invention has a 3′ terminal nucleotide, or

preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide primers hybridizing to either the coding or noncoding strand are contemplated by the invention. A preferred ASO primer for detecting NPY gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CTTGCATATTCATTC (SEQ ID NO:20); GTTAAACCTGTTGAA (SEQ ID NO:21);
CTTGCATATTCATGC (SEQ ID NO:22); GTTAAACCTGTTGCA (SEQ ID NO:23);
GGCACCCAGTGGGTT (SEQ ID NO:24); AACAGGACTACCAAC (SEQ ID NO:25);
GGCACCCAGTGGGCT (SEQ ID NO:26); AACAGGACTACCAGC (SEQ ID NO:27);
CTGGGGCGGGAGGGT (SEQ ID NO:28); ACCCACACCCCAACC (SEQ ID NO:29);
CTGGGGCGGGAGGTT (SEQ ID NO:30); ACCCACACCCCAAAC (SEQ ID NO:31);
CGGTTCCAGCCCCCT (SEQ ID NO:32); AGTGGCGGGGGGAGG (SEQ ID NO:33);
CGGTTCCAGCCCCAT (SEQ ID NO:34); AGTGGCGGGGGGATG (SEQ ID NO:35);
CCCGCTCTGTGGAGC (SEQ ID NO:36); CAGGGCCCCAGGGCT (SEQ ID NO:37);
CCCGCTCTGTGGATC (SEQ ID NO:38); CAGGGCCCCAGGGAT (SEQ ID NO:39);
GAAGGGAGAAAAGTG (SEQ ID NO:40); TTCCTGCTGGGTCAC (SEQ ID NO:41);
GAAGGGAGAAAAGCG (SEQ ID NO:42); TTCCTGCTGGGTCGC (SEQ ID NO:43);
GGACGGCGCCCGGAG (SEQ ID NO:44); CCACCTTGCGGGCTC (SEQ ID NO:45);
GGACGGCGCCCGGGG (SEQ ID NO:46); CCACCTTGCGGGCCC (SEQ ID NO:47);
GCTCTGAATCCCCAA (SEQ ID NO:48); CAACGGACGGGCTTG (SEQ ID NO:49);
GCTCTGAATCCCCCA (SEQ ID NO:50); and CAACGGACGGGCTGG (SEQ ID NO:51).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site. A particularly preferred oligonucleotide primer for detecting NPY gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCATATTCAT (SEQ ID NO:52); AAACCTGTTG (SEQ ID NO:53);
ACCCAGTGGG (SEQ ID NO:54); AGGACTACCA (SEQ ID NO:55);
GGGCGGGAGG (SEQ ID NO:56); CACACCCCAA (SEQ ID NO:57);
TTCCAGCCCC (SEQ ID NO:58); GGCGGGGGGA (SEQ ID NO:59);
GCTCTGTGGA (SEQ ID NO:60); GGCCCCAGGG (SEQ ID NO:61);
GGGAGAAAAG (SEQ ID NO:62); CTGCTGGGTC (SEQ ID NO:63);
CGGCGCCCGG (SEQ ID NO:64); CCTTGCGGGC (SEQ ID NO:65);
CTGAATCCCC (SEQ ID NO:66); and CGGACGGGCT (SEQ ID NO:67).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic

sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

NPY genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized NPY genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the NPY gene in an individual. As used herein, the terms "NPY genotype" and "NPY haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the NPY gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid mixture comprising the two copies of the NPY gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more of the polymorphic sites selected from PS1-PS5, PS9, PS13 and PS14 in the two copies to assign a NPY genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair atone or more of the polymorphic sites selected from the group consisting of PS6-PS8, PS10-PS12 and PS15-PS19 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-19.

Typically, the nucleic acid mixture is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample

must be obtained from an organ in which the NPY gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ nontranscribed regions. If a NPY gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid molecule containing only one of the two copies of the NPY gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more of the polymorphic sites PS1-PS5, PS9, PS13 and PS14 in that copy to assign a NPY haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the NPY gene or fragment such as one of the methods described above for preparing NPY isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two NPY gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional NPY clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the NPY gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at one or more of the polymorphic sites PS6-PS8, PS10-PS12 and PS15-PS19. In a particularly preferred embodiment, the nucleotide at each of PS1-19 is identified.

In a preferred embodiment, a NPY haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more of the polymorphic sites selected from PS1-PS5, PS9, PS13 and PS14 in each copy of the NPY gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-19 in each copy of the NPY gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the NPY gene, or fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in

individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping a polymorphic site not disclosed herein that is in linkage disequilibrium with the polymorphic site that is of interest. Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stevens, JC 1999, *Mol. Diag.* 4: 309-17). Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., Proc. Natl. Acad. Sci. USA 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., Science 241:1077-1080, 1988). Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant.

In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the NPY gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., Proc. Natl. Acad. Sci. USA 82:7575, 1985; Meyers et al., Science 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. Ann. Rev. Genet. 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., Genomics 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., Nucl. Acids Res. 18:2699-2706, 1990; Sheffield et al., Proc. Natl. Acad. Sci. USA 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., Nucl. Acids Res. 17:8392, 1989; Ruaño et al., Nucl. Acids Res. 19, 6877-6882, 1991; WO 93/22456; Turki et al., J.

Clin. Invest. 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In another aspect of the invention, an individual's NPY haplotype pair is predicted from its NPY genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a NPY genotype for the individual at two or more polymorphic sites selected from PS1-PS5, PS9, PS13 and PS14, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing NPY haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the NPY haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_i / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping

method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996). A preferred process for predicting NPY haplotype pairs from NPY genotypes is described in copending U.S. Provisional Application Serial No. 60/198,340.

In one embodiment of this method for predicting a NPY haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

The invention also provides a method for determining the frequency of a NPY genotype or NPY haplotype in a population. The method comprises determining the genotype or the haplotype pair for the NPY gene that is present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of the polymorphic sites PS1-PS5, PS9, PS13 and PS14 in the NPY gene; and calculating the frequency any particular genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment). A preferred process for predicting NPY haplotype pairs from NPY genotypes is described in pending U.S. Provisional Application Serial No. 60/198,340.

In another aspect of the invention, frequency data for NPY genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a NPY genotype or a NPY haplotype. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s) or haplotype(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by

accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s) or haplotype(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the NPY gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that NPY genotype or haplotype. Preferably, the NPY genotype or haplotype being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, respectively, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting NPY or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a NPY genotype or haplotype, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease

susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the NPY gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and NPY genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their NPY genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the NPY gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between NPY haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., supra Ch. 10), or other global or local optimization approaches (see discussion in Judson, supra) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the NPY gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, supra, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of NPY genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the NPY gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the NPY gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying NPY genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the NPY gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The NPY polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and

Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the NPY gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the NPY gene were amplified using the PCR primer pairs listed below, with the sequences presented in the 5′ to 3′ direction and nucleotide positions shown for each region corresponding to the reverse complement of the indicated GenBank Accession No.

Accession No. ACC004485.1

Fragment 1		
Forward primer: 154135-154158	5'- GGAGCATTCATTCACGGAACTTTC -3'	(SEQ ID NO:68)
Reverse primer: Complement of 154540-154520	5'- GGCTCGGTCGAGACATGAGTG -3'	(SEQ ID NO:69)
PCR Product 406 nt		
Fragment 2		
Forward primer: 154364-154387	5'- CATTTAACCACCGTCACTTTGGAC -3'	(SEQ ID NO:70)
Reverse primer: Complement of 154901-154879	5'- AGTCCTGCCAGAGATAGGAGCAG -3'	(SEQ ID NO:71)
PCR Product 538 nt		
Fragment 3		
Forward primer: 154726-154747	5'- CCAGTTGCCTCACTCCAACAGC -3'	(SEQ ID NO:72)
Reverse primer: Complement of 155161-155142	5'- TGACTCCCACCCGCCACTTC -3'	(SEQ ID NO:73)
PCR Product 436 nt		
Fragment 4		
Forward primer: 154967-154987	5'- ACTTGGGCTTTGGTGTTGTCC -3'	(SEQ ID NO:74)
Reverse primer: Complement of 155532-155511	5'- AATCAGAGGGGAGCAGAGGAAC -3'	(SEQ ID NO:75)
PCR Product 566 nt		
Fragment 5		
Forward primer: 155377-155399	5'- GAACTTGCTAGAGACGCAGCCTC -3'	(SEQ ID NO:76)
Reverse primer: Complement of 155881-155861	5'- CTTCAGGGGATCAACGCTGAC -3'	(SEQ ID NO:77)
PCR Product 505 nt		

Fragment 6
Forward primer:
155760-155780 5'- TGGATTCTTGGGCTCCAAATC -3' (SEQ ID NO:78)
Reverse primer:
Complement of 156262-156243 5'- GCTTGGGGATTCAGAGCACC -3' (SEQ ID NO:79)
PCR Product 503 nt

Fragment 7
Forward primer:
156031-156053 5'- GCTTCTTGGTCCCTGAGACTTCG -3' (SEQ ID NO:80)
Reverse primer:
Complement of 156556-156536 5'- CCCTTTCCCTAACATCCCCAG -3' (SEQ ID NO:81)
PCR Product 526 nt

Fragment 8
Forward primer:
160417-160442 5'- GCTTGCTTGTTACAGATGAACACCTG -3' (SEQ ID NO:82)
Reverse primer:
Complement of 160870-160848 5'- CGTCAAAGGAGCACCAAATAACC -3' (SEQ ID NO:83)
PCR Product 454 nt

Fragment 9
Forward primer:
162593-162616 5'- CAGGAAACTTTTCAACAGTTCCCG -3' (SEQ ID NO:84)
Reverse primer:
Complement of 163029-163009 5'- GCGAAACGAACCCTGAATCTG -3' (SEQ ID NO:85)
PCR Product 437 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 20 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 2 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	=15.6µl

Amplification profile:
94°C - 2 min. 1 cycle

94°C - 30 sec.
70°C - 45 sec.
72°C - 1 min. } 10 cycles

94°C - 30 sec.
64°C - 45 sec.
72°C - 1 min. } 35 cycles

Sequencing of PCR Products

The PCR products were purified by Solid Phase Reversible Immobilization using the protocol developed by the Whitehead Genome Center. A detailed protocol can be found at http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI_pcr.html.

Briefly, five µl of carboxyl coated magnetic beads (10 mg/ml) and 60 µl of HYB BUFFER (2.5M NaCl/20% PEG 8000) were added to each PCR reaction mixture (20 µl). The reaction mixture was mixed well and incubated at room temperature (RT) for 10 min. The microtitre plate was placed on a magnet for 2 min and the beads washed twice with 150 µl of 70% EtOH. The beads were air dried for 2 min and the DNA was eluted in 25 µl of distilled water and incubated at RT for 5 min. The beads were magnetically separated and the supernatant removed for testing and sequencing.

The purified PCR products were sequenced in both directions using the primer sets described previously or those listed below, in the 5′ to 3′ direction.

Fragment 7
Reverse sequencing primer:
Complement of 156535-156515 5'- GATCTCCTGGTGTGCAGGCAC -3' (SEQ ID NO:86)

Fragment 9
Forward sequencing primer
162638-162661 5'- AAATGTCTCACCCTTGCTCATACC -3' (SEQ ID NO:87)
Reverse sequencing primer
Complement of 163023-163000 5'- CGAACCCTGAATCTGCTAGTTGAG -3' (SEQ ID NO:88)

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., Nucleic Acids Res. 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the NPY gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the NPY Gene

Polymorphic Site Number	PolyId	Nucleotide Position in GenBank	Nucleotide Position in Fig 1	Reference Allele	Variant Allele
PS1	98214	31293(Acc#AC004485.1)	154224	T	G
PS2	98218	30676(Acc#AC004485.1)	154841	T	C
PS3	98219	30454(Acc#AC004485.1)	155063	G	T
PS4	98220	30398(Acc#AC004485.1)	155119	C	A
PS5	98222	30105(Acc#AC004485.1)	155412	G	T
PS6 [R]	98223	30104(Acc#AC004485.1)	155413	C	A
PS7 [R]	98224	30097(Acc#AC004485.1)	155420	G	T
PS8 [R]	98227	29955(Acc#AC004485.1)	155562	G	A
PS9	98230	29563(Acc#AC004485.1)	155954	T	C
PS10 [R]	98231	29498(Acc#AC004485.1)	156019	C	T
PS11 [R]	98232	29478(Acc#AC004485.1)	156039	G	A
PS12 [R]	98233	29416(Acc#AC004485.1)	156101	A	G
PS13	98234	29400(Acc#AC004485.1)	156117	A	G
PS14	98235	29258(Acc#AC004485.1)	156259	A	C
PS15 [R]	98236	29207(Acc#AC004485.1)	156310	T	C
PS16 [R]	98237	29143(Acc#AC004485.1)	156374	G	A
PS17 [R]	98238	22735(Acc#AC004485.1)	162782	A	G
PS18 [R]	98239	22600(Acc#AC004485.1)	162917	A	G
PS19 [R]	98240	22569(Acc#AC004485.1)	162948	T	C

[R] Reported previously in literature

EXAMPLE 2

This example illustrates analysis of the NPY polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4(Part1). Genotypes and Haplotype Pairs Observed for NPY Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP Pair	
1	T	T	G	C	G	C	G	G	T	C	9	9
2	T	C	G	C	G	C	T	G	T	C	4	4
3	T	T	G	C	G	C	G	G	T	T	13	13
4	T	C/T	G	C	G	C	T/G	G	T	C/T	4	14
5	T	C	G	C	G	C	T	G	T	C	4	6
6	T	T	G	C	G	C	G	G	T	T	13	14
7	T	C/T	G	C	G	C	T/G	G	T	C/T	4	17
8	T	C/T	G	C	G/T	C	T/G	G	T	C/T	4	19
9	T	C/T	G	C	G	C	T/G	G	T	C	4	11
10	T	T	G	C	G	C	G	G	T	T	13	16
11	T	C/T	G	C	G	C	T/G	G	T	C/T	4	15
12	T	T	G	C	G	C/A	G	G	T	T	13	7
13	T	C/T	G	C	G	C	T	G	T/C	C	4	18
14	T	C	G	C	G	C	T	G/A	T	C	4	3
15	T	C	G	C	G	C	T	G	T	C	4	5
16	T	T	G	C	G	C	G/T	G	T/C	T/C	13	18
17	T	C/T	G/T	C	G	C	T/G	G	T	C/T	4	20
18	T	C	G	C/A	G	C	T	G	T	C	4	2
19	T	C/T	G	C	G	C	T/G	G	T	C/T	4	12
20	T	T	G	C	G	C	G	G	T	T/C	13	9
21	T	C/T	G	C	G	C	T/G	G	T	C	4	9
22	T	T	G	C	G	C	G	G	T	T/C	13	8
23	T	T	G	C	G	C	G	G	T	T/C	13	10
24	T	C/T	G	C	G	C	T/G	G	T	C	4	8
25	T	T	G	C	G	C	G	G	T	C	9	8
26	T	C/T	G	C	G	C	T/G	G	T	C/T	4	13
27	T/G	T/C	G	C	G	C	G	G	T	C/T	9	1

Table 4(Part2). Genotypes and Haplotype Pairs Observed for NPY Gene

Genotype Number	Polymorphic Sites PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	HAP Pair	
1	G	A	A	A	T	G	A	A	T	9	9
2	G	A	A	A	T	G	A	A	T	4	4
3	A	G	A	A	T	G	A	A	T	13	13
4	G/A	A/G	A	A	T	G	A	A/G	T	4	14
5	G	A	A/G	A	T	G	A	A	T	4	6
6	A	G	A	A	T	G	A	A/G	T	13	14
7	G	A/G	A	A	T	G/A	A/G	A	T	4	17
8	G/A	A/G	A	A	T	G	A	A	T/C	4	19
9	G	A	A	A	T	G	A/G	A	T	4	11
10	A/G	G	A	A	T	G/A	A	A	T/C	13	16
11	G/A	A/G	A	A	T	G	A/G	A	T	4	15
12	A	G	A	A	T	G	A	A	T	13	7
13	G	A	A	A	T	G	A	A	T	4	18
14	G	A	A	A	T	G	A	A	T	4	3
15	G	A	A	A/C	T	G	A	A	T	4	5
16	A/G	G/A	A	A	T	G	A	A	T	13	18
17	G/A	A/G	A	A	T	G	A	A	T	4	20
18	G	A	A	A	T	G	A	A	T	4	2
19	G/A	A/G	A	A	T	G	A	A	T/C	4	12
20	A/G	G/A	A	A	T	G	A	A	T	13	9
21	G	A	A	A	T	G	A	A	T	4	9
22	A/G	G/A	A	A	T/C	G	A	A	T	13	8
23	A/G	G/A	A	A	T	G	A	A/G	T	13	10
24	G	A	A	A	T/C	G	A	A	T	4	8
25	G	A	A	A	T/C	G	A	A	T	9	8
26	G/A	A/G	A	A	T	G	A	A	T	4	13
27	G	A/G	A	A	T	G	A	A	T	9	1

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using an extension of Clark's algorithm (Clark, A.G. (1990) *Mol Bio Evol* 7, 111-122), as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms". In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. In our analysis, the list of haplotypes is augmented with haplotypes obtained from a three-generation Caucasian family and a two-generation African-American family. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 20 human NPY haplotypes shown in Table 5 below.

Table 5. Haplotypes Identified in the NPY Gene

Haplotype	Polymorphic Sites																	
	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS	PS
Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	G	C	G	C	G	C	G	G	T	T	G	G	A	A	T	G	A	A	T
2	T	C	G	A	G	C	T	G	T	C	G	A	A	A	T	G	A	A	T
3	T	C	G	C	G	C	T	A	T	C	G	A	A	A	T	G	A	A	T
4	T	C	G	C	G	C	T	G	T	C	G	A	A	A	T	G	A	A	T
5	T	C	G	C	G	C	T	G	T	C	G	A	A	C	T	G	A	A	T
6	T	C	G	C	G	C	T	G	T	C	G	A	G	A	T	G	A	A	T
7	T	T	G	C	G	A	G	G	T	T	A	G	A	A	T	G	A	A	T
8	T	T	G	C	G	C	G	G	T	C	G	A	A	A	C	G	A	A	T
9	T	T	G	C	G	C	G	G	T	C	G	A	A	A	T	G	A	A	T
10	T	T	G	C	G	C	G	G	T	C	G	A	A	A	T	G	A	G	T
11	T	T	G	C	G	C	G	G	T	C	G	A	A	A	T	G	G	A	T
12	T	T	G	C	G	C	G	G	T	T	A	G	A	A	T	G	A	A	C
13	T	T	G	C	G	C	G	G	T	T	A	G	A	A	T	G	A	A	T
14	T	T	G	C	G	C	G	G	T	T	A	G	A	A	T	G	A	G	T
15	T	T	G	C	G	C	G	G	T	T	A	G	A	A	T	G	G	A	T
16	T	T	G	C	G	C	G	G	T	T	G	G	A	A	T	A	A	A	C
17	T	T	G	C	G	C	G	G	T	T	G	G	A	A	T	A	G	A	T
18	T	T	G	C	G	C	T	G	C	C	G	A	A	A	T	G	A	A	T
19	T	T	G	C	T	C	G	G	T	T	A	G	A	A	T	G	A	A	C
20	T	T	T	C	G	C	G	G	T	T	A	G	A	A	T	G	A	A	T

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for genotyping the neuropeptide Y (NPY) gene of an individual, comprising determining for the two copies of the NPY gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1-PS5, PS9, PS13 and PS14.

2. The method of claim 1, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the NPY gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing at least one of the polymorphic sites;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

3. A method for haplotyping the neuropeptide Y (NPY) gene of an individual which comprises determining, for one copy of the NPY gene present in the individual, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1-PS5, PS9, PS13 and PS14.

4. The method of claim 3, wherein the determining step comprises
   (a) isolating from the individual a nucleic acid molecule containing only one of the two copies of the NPY gene, or a fragment thereof, that is present in the individual;
   (b) amplifying from the nucleic acid molecule a target region containing at least one of the polymorphic sites;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

5. A method for predicting a haplotype pair for the neuropeptide Y (NPY) gene of an individual comprising:
   (a) identifying an NPY genotype for the individual at two or more of the polymorphic sites PS1-PS5, PS9, PS13 and PS14;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) accessing data containing the NPY haplotype pairs determined in a reference population; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

6. A method for identifying an association between a trait and at least one genotype or haplotype of the neuropeptide Y (NPY) gene which comprises comparing the frequency of the genotype or haplotype in a population exhibiting the trait with the frequency of the genotype or haplotype in a reference population, wherein the genotype or haplotype comprises a nucleotide pair or nucleotide located at one or more polymorphic sites selected from the group consisting of PS1-PS5, PS9, PS13 and PS14, wherein a higher frequency of the genotype or haplotype in the trait population than in the reference population indicates the trait is associated with the genotype or haplotype.

7 The method of claim 6, wherein the haplotype is selected from haplotype numbers 1-20 shown in Table 5.

8. The method of claim 7, wherein the trait is a clinical response to a drug targeting NPY.

9. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the neuropeptide Y (NPY) gene at a polymorphic site selected from PS1-PS5, PS9, PS13 and PS14.

10. The composition of claim 9, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the NPY gene at a region containing the polymorphic site.

11. The composition of claim 10, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of of SEQ ID NOS:4-19, the complements of SEQ ID NOS: 4-19, and SEQ ID NOS:20-51.

12. The composition of claim 9, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

13. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for neuropeptide Y (NPY) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, thymine at PS5, cytosine at PS9, guanine at PS13 and cytosine at PS14; and

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

14. The isolated polynucleotide of claim 13 which comprises a NPY isogene.

15. The isolated polynucleotide of claim 13 which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

16. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 13, wherein the organism expresses a NPY protein encoded by the first nucleotide sequence.

17. The recombinant organism of claim 16 which is a nonhuman transgenic animal.

18. The isolated polynucleotide of claim 13, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the NPY gene, the fragment comprising one or more polymorphisms selected from the group consisting of guanine at PS1, cytosine at PS2, thymine at PS3, adenine at PS4, thymine at PS5, cytosine at PS9, guanine at PS13 and cytosine at PS14.

19. A computer system for storing and analyzing polymorphism data for the neuropeptide Y gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.

20. A genome anthology for the neuropeptide Y (NPY) gene which comprises NPY isogenes defined by any one of haplotypes 1-20 shown in Table 5.

POLYMORPHISMS IN THE NPY GENE

GAGAGAAGAG TGGACACGTT TAGATACGGA AGATATAAAA AAGATTCATT
GCTTTATATT GTCCTGTAAC TTAAAAAAAA GATTCAAGTT GAAACTCTAG 152300
AGATGAAAGC TACCAGGTAT AAGGTGCAAA ATACACTGGA TAGGATTGGC
AGTAATCAGA TATTACAGAA GAAAACATTA GTGAACTCAC AGACATAGCA 152400
ATAGAGACTA TCCAAAATAA AATATGAAGG AAAAAATGAG CAAAACAGAA
GCAAAAACAG ATCACCAACA AACTGTGGGC CAATATACAC TCACCGTCTG 152500
ACTCAGCCAT TCTGGTTTTA AGTATTTTCC AAGAGAATTA AAGGCTTTAT
GTCCATACAA AGACTTGCAT ATGAATGTCA AGGCAGCTTT ATTCGTAAAA 152600
GTTTCAAGTG AAAATAACCA AAAATTCATG AGCAGTAGAA TTGATTAAAA
ATTTTGAAGT ATATCATATA ATGAAATACT GCCCAGCAAT AAAAAAGAAT 152700
AAACTATTGC TACATTTAAC AATGTGAGTG GATTCAAAAA TAATGATGCT
AAGTGAAAGA AGTCAGACAA AGAGTACCTA CCGTGGAACT ATATTTATGT 152800
GCAATTCTAG AAAACGCAAA CAAATCTCTA GTGATAGCAG ATCAATGGAG
GAGGGGGTGA GGAAGAGGCA AAAGGAAGGA ATTACAAAGG TACATGTGGA 152900
AATTTGGGGA AGTGATGTAT AATCTTCATT TTCTTGGTTT CAGTGATGGC
TTCCTGAGCA TATAACACAC ATGCATATGT GTAAGAATAT ATCCAATTGA 153000
ATATTTTACA CATTTGTAGC TTTGTATATA TTCAGCAGTA TCTATTCATT
CCTTCTGCTT CTTTACCTTA GGTTTAGTCA TGAAACCACT GGGTTTCATG 153100
TGGTCCAGGT TCTATCTGAA TCAAGCTATG AACCTGCTAT CAGTAAGACC
TCCTAATTGC AAACCACATG GACCTTTACC GGTTATTTGA CTTCTCTCGG 153200
TCATTACCTT CTCTAGTCTC CACTGATTCA CCTCCTTTTG AATCTCAACA
AGCTCAACAG TATCTCAATG TTATCTTAAC AAAGCTGTTA GAAATAAACA 153300
ACTTGACTTT GAAACAAAGG AGAGAAATGG CTCAGCAGCT AGGAGAAGAT
GCAGGGTTGA GTTGCCAACA GTTGGTTTTG TTTGTATTTT GGCCAGGGGA 153400
TGTGGCTTGG ACTGGAGAGA AAGGAGATAA GGATGTAAGC ACATGTAGGG
CATATCACCC CCTATTTTTT ATTCTCTGAA TCCTTAACCC TCAGAATAAG 153500
TTCTTATTCT TGAGAATCAA TGACATTATC TTAAGCTAAA TTAATCAAGC
CTCCACAGTG TTCTTCTCTC AATAGTGGTG TGGGCCTTCC TAGAAGTAAT 153600
TTTTCCCAAA TTCAGTGATA CATTTTAAGT TCAGATTTTA ATTGATATGA
ATCTGTGATA CACTCTAAAA TAAGATTATT TTATTGAAAA GTGGACTGTA 153700
ACTTTCCCTT TATCTAGGAA GAGCTCTAAG TTAGAAGATG TTTTGCACTT
TTACCGAAGG CTGTGTCTTG TAAGCACCCC CGAGCAACTC TGAGAGCCTT 153800
GATTTTTGTG TCCTCAGCAT ATGTTTGTGT AATACAGAAA GAGAAGCAGT
TGCCAAGTGA AAGGGATGTT GGTCTCCAAA ATTATAGTTT GATCCCACAA 153900
ACACACAAAC ACATACATGC AAAGGATTGT TTGCTTCACG GTTTTTGATA
TTTAATTCAA TGCTGTTGGA ACAGCACAAA AACTAAGTGT CAGTTTAACA 154000
GAATCACTTG TCCTTTTAGC ATTAAAATAA CATGGAACTT AATGCTTTAA
TTTCCCAACA TGCCTTTTTA TTTAGAAAGA TTCAGACTTA TATTTCATTT 154100
AGAAATAAAA TGCCATTTTA TTTAGAAAGA TACAGGAGCA TTCATTCACG
GAACTTTCAG ATCTCAGTCC ACTGCATAAA ATCTTGATCC TGTAATAATA 154200
GTTTCTGTAT CTTGCATATT CATTCAACAG GTTTAACGCG ATGAGCAAAT
G
TAATGTTCAT CGTTTTTAAC ATGTTTCATC TTAATCAGAA CCCACATTCT 154300
CAACGTTAAT TGAACGTACA TAGGACTATA CAAGGGTTAG TAAATAAGAC
AGAAACTGTT GTTCATTTAA CCACCGTCAC TTTGGACCAA AAAAGAAAAA 154400
ATATATATTT TTAAAATTGA GCTTAAAAGA GTCTCTAGAA GCTGGAAGCG
TGGCTCTTTT TCAGCAAACT GGGGGAATAG GTTTACCGTG TTCCCCCTCT 154500
GGGGAATTTT GAGTCGCCAC ACTCATGTCT CGACCGAGCC TGGCTCGCTG
CGTCTGAGCG AGTACTTGAG GAAGGCTGAT CTAGAAAAAC CAGCTGAGAG 154600
AAGGGGCAGA AGCCCCTGAA ACCACGGGCG GGGGTGGGGT GGGGAGCGCA

FIGURE 1A

```
GCTTTGGGAC CCTCTAGCCG GAGACTTCCG GCAGCTGCCT CCGACTTGTT      154700
CTAAGTACAG GAAAAATCTG TGCGCCCAGT TGCCTCACTC CAACAGCGCG
CAGTTGTGCC CGGCGAGGAT GCCGCGCTAG TCGTGGAGAT GCCCCACCAC      154800

AAAGAGGATT CAGGTGCTTC CTACTCCGGC ACCCAGTGGG TTGGTAGTCC
                                             C
TGTTGGCAGG AGACAAGAAT CGTCTGGGCT GCTCCTATCT CTGGCAGGAC      154900
TAGACGGGGC GTGAAGGAAA GAAGGAAAGA AGGAAAGCAG GGATCGGGCA
CTGCCCGAGG GCAGATACTT GGGCTTTGGT GTTGTCCAGC GCGCTCGGAG      155000
TGCGCTGCCT CGCTCACGCG GTCCCAGGCC CCGCTTCTTC AGGCAGTGCC
TGGGGCGGGA GGGTTGGGGT GTGGGTGGCT CCCTAAGTCG ACACTCGTGC      155100
            T
GGCTGCGGTT CCAGCCCCCT CCCCCCGCCA CTCAGGGGCG GGAAGTGGCG
                  A
GGTGGGAGTC ACCCAAGCGT GACTGCCCGA GGCCCCTCCT GCCGCGGCGA      155200
GGAAGCTCCA TAAAAGCCCT GTCGCGACCC GCTCTCTGCA CCCCATCCGC
TGGCTCTCAC CCCTCGGAGA CGCTCGCCCG ACAGCATAGT ACTTGCCGCC      155300
CAGCCACGCC CGCGCGCCAG CCACCGTGAG TGCTACGACC CGTCTGTCTA
GGGGTGGGAG CGAACGGGGC GCCCGCGAAC TTGCTAGAGA CGCAGCCTCC      155400
CGCTCTGTGG AGCCCTGGGG CCCTGGGATG ATCGCGCTCC ACTCCCCAGC
           TA     T
GGACTATGCC GGCTCCGCGC CCCGACGCGG ACCAGCCCTC TTGGCGGCTA      155500
AATTCCACTT GTTCCTCTGC TCCCCTCTGA TTGTCCACGG CCCTTCTCCC
GGGCCCTTCC CGCTGGGCGG TTCTTCTGAG TTACCTTTTA GCAGATATGG      155600
           A
AGGGAGAACC CGGGACCGCT ATCCCAAGGC AGCTGGCGGT CTCCCTGCGG
GTCGCCGCCT TGAGGCCCAG GAAGCGGTGC GCGGTAGGAA GGTTTCCCCG      155700
GCAGCGCCAT CGAGTGAGGA ATCCCTGGAG CTCTAGAGCC CCGCGCCCTG
CCACCTCCCT GGATTCTTGG GCTCCAAATC TCTTTGGAGC AATTCTGGCC      155800
CAGGGAGCAA TTCTCTTTCC CCTTCCCCAC CGCAGTCGTC ACCCCGAGGT
GATCTCTGCT GTCAGCGTTG ATCCCCTGAA GCTAGGCAGA CCAGAAGTAA      155900
CAGAGAAGAA ACTTTTCTTC CCAGACAAGA GTTTGGGCAA GAAGGGAGAA
AAGTGACCCA GCAGGAAGAA CTTCCAATTC GGTTTTGAAT GCTAAACTGG      156000
  C
CGGGGCCCCC ACCTTGCACT CTCGCCGCGC GCTTCTTGGT CCCTGAGACT
                T                    A
TCGAACGAAG TTGCGCGAAG TTTTCAGGTG GAGCAGAGGG GCAGGTCCCG      156100
ACCGGACGGC GCCCGGAGCC CGCAAGGTGG TGCTAGCCAC TCCTGGGTTC
G           G
TCTCTGCGGG ACTGGGACGA GAGCGGATTG GGGGTCGCGT GTGGTAGCAG      156200
GAGGAGGAGC GCGGGGGGCA GAGGAGGGAG GTGCTGCGCG TGGGTGCTCT
GAATCCCCAA GCCCGTCCGT TGAGCCTTCT GTGCCTGCAG ATGCTAGGTA      156300
    C
     [EXON 1: 156291..
ACAAGCGACT GGGGCTGTCC GGACTGACCC TCGCCCTGTC CCTGCTCGTG
        C
TGCCTGGGTG CGCTGGCCGA GGCGTACCCC TCCAAGCCGG ACAACCCGGG      156400
                  A
CGAGGACGCA CCAGCGGAGG ACATGGCCAG ATACTACTCG GCGCTGCGAC
ACTACATCAA CCTCATCACC AGGCAGAGGT GGGTGGGACC GCGGGACCGA      156500
               ..156478]
TTCCGGGAGC GCCAGTGCCT GCACACCAGG AGATCCTGGG GATGTTAGGG
```

FIGURE 1B

```
AAAGGGATTG TTTCTTTTCC TTCGCTCTAT CCCAGGGCAG GACAGTATCA      156600
GGCACTTAGT CAGCTCTAGG TAAATGTTTG TACAGGGCAC ACTCTACACA
AAATGGGTAC CTTCCATTTT GTGCAACTAC AGTCACAGAG TCGTGATCCC      156700
CAGATTCAGG TTCCCCAGGC TGGTAGGCTG GCAATCTCCT CTCACTCACC
TCTTATGGTT TGTTGTGGTT CTTACGGCAG TGGGGCCCGG TCCAGAAATC      156800
TCGAAAGTAC CCAGTGAAAG GGGCAAGAAT GCGCCAGAGA AATGCTGTAG
GGGGAAACGC TAGCAAGGTG TCTAGGAGAA ACAGAACGAC CACCAAAGAA      156900
AACCAAACCA AGGAGTAAAC TGCAGGGTTG CCACAGACAT TGTCAGACTT
TCCGGCCTGC CCAGGGCTAA TTGAATGACA GTATTTAGAA AAAGACAAAT      157000
AGAGCTATTT TCTTTTCTTC TTCACTGTGT TTTTTTTGCA GTCTCACTCC
CATATTTGAA AATTCTAGTC CCAATAATTA GAGATTTAGC CTGAAAACAG      157100
TTTATCTAAA ATGCAGAGGT CTCCCTATTT TATAGCTGGT GAATTCTGAA
AAGAAAGTGC CCACTTGGTA TTTCATTAAG AAGAAACTTA CGTTGGTCCA      157200
GTAACTTTAG GATGAATTCA GGGATTACAT GGAATCTGAG AAATTGTAAA
CTTTATTCTG AACATTCCGA AAGTCAAAAT GAAATGCCAT ATTTTCACTC      157300
TCAACATTGA CAATTAAGGG AAAAATCAAA TCCAAACTAT TTTACAAATT
ACATAGGATT TAAATTTTGT TTTCAAAATT GGTTGATCTT AGTTCTAAAT      157400
CATTTAATAT CTTATTTTAA TTACTGAGAT TAAATGATTT TAGTCATTTG
ACCGTATCTA AAGGATAAGC CCTTTAGAAG CATTAATGTT AGTTAACATG      157500
TTTAAATTAT ATGTTTAAAG TATTGTGGGG TTTATATATT TATTTATTTA
GTCCATATTT TTACTGTTTA CAAGTCAGGG AGTAGTTGGA AAGGATCCAA      157600
AAGGGGGGAA GATTTATACT TTTAAAATAA AGTTTGCTTT TTCCTGTTTC
ATCAAAAAGG TATTTAACAT CCTAAAACAC TTTTGAAAAA GCCCCAATTT      157700
ATATAGAATC TGAAAAACAT GAAAATGCCA GAATTAAATA GTAGTTGTTT
GTGATCCCTT GAAATCTGAT CATTTTAATT AGCAAAAAGC AATAATATTT      157800
TATATTCCAT ACTTCCCAGA AAGTTAATTG ATTCAAATTC TTGCTATGTT
ATGCCGTCAT GCAGATCAAA GAGAACTTGT ACAAACAAAA AGATATTAAC      157900
TGTTTTTTCT AGAAAAACCT TTGTTAGTTA TTTAGCTGTA GCATAATACA
CAGTAGAACT CATTTTTCCG AAGATGATTG CCATAAAGCT TTACCATTTT      158000
ACAAAGCTCT AGCATTTTAA AATGTTTTCC TAGGCATTCT CTCATTCAGT
CGTTACCTCA AAGCTTTGAA GGGAGTATAA TTATCTTCAT TTCACAAATC      158100
ATGAAGCTAA TGACATTTCA TGGAAGCTAT TTTTATAGGA ATTATGGAAT
TGGAGCCCTC CTTTGAGTGT CTTTGGATGT TGTCACCACC ATCAACAGGC      158200
AGTGTTATCT CCAACATGTT TGATGAGAAA AAAACATAAA GGAAAGGGAC
TTGGGTTTCA TGGCAAGAGA TTCAGCTTAG ATGGGAATGA GCTGGTCTCT      158300
GAGGCAGAGT TTCCTGTCTC ATCTTTGTAT AAACTCCTTC TTAGAAACCA
GCAGAGGCTC TTAATGTACG TCCATCTTCT CCTTCCCACC ACAGACACCC      158400
ACTCCTATGT GCGTTTCTGA AAATTACAGG GTCATTTGTG GATGTGCTTT
AATTTAGTAT AGATATAATC TGATATCATT ATCTTTCTTT CCCCCCGCCC      158500
CCCGCCACCA GCATTCCCCC TGCACTCAAT TTGTGCTTAT ATTCCTTTTG
CTGTATATAT ACCCCTGGCT CATTGCTTCT GGAAGCTGCA TTACTTATTC      158600
TAGAGTGTGG GTCTCCTACA ATTTATTTAT ATAGCCCCCT AAATCAGTGG
TTCTCAACTG GGGGTGATTT TATGCTCCCC AGGTGTGATT TGGCAGTGTC      158700
TGGAGACATT TTTGGTTTTC ACAACTGGAG GTGGGCACTG CTGGTACCTA
GGTGTATAGA TGCCAAGGAG GCTGCTAAAT ATCCTACCAT GCCCAGGACA      158800
GCCCCCATTA CACAGAATAG CCAGTCCCAG AGAAACTCAG CCTTAGATTA
GTGTTTCTCC TTCCTGGCTG CACATTAAAA TTATCAGGGG AGCTTTTAAA      158900
ATACACTGAT GCTGGGATCC ACCCCTGACC AATTAAATCA GACCCTCTGG
GATGGGGCCT GGGCTTTAGA GTGTGTTTTT AACACTCCCC AGGTGATTCT      159000
AATGTGCAGC CAGGTTTGAG AATCTTTACT CTAAGGACTC AATCTTTGAA
AGTTACAGCA TTGTAGAGGG GAAGGATGAT CCAGAATTAA TTAAAGAAGG      159100
AAGAAGAGTG AACAGTACTG AACTGGATGC AAGATTATGA CTGTGCAAAT
```

FIGURE 1C

```
CCTCCTATTT CAGATTTAGG TTCTTTACTG GTAAATTGGA TAAAAAGTTG      159200
GCAGTTCTTT TACTCACTGG AGCATGATGA AGATAAATGA GTTACAGTTG
CCCTAGACTT TTACGATTAG TAAAACCAAT GGATGGACAG AATAAATAAG      159300
TATAAGCAGA CACCTATAGT CTAATATCGC AACTCCAGAA AATCTTTGAG
AAAATCTCAC ACATGATGGG AGTGGGACAG GGAGTTGTTT TGTTTTTGCT      159400
TCCCTTGAGT CTCAGCAAAA GAACAGTCAT AGAAAATATT TCCAAAAAAT
TTAGATATAA TTTTGCTTGG CAATTAGAAT ATACCCTCTG TCAGCTTTGT      159500
TCAGTTGTTG GACCTATTAA CCAATCCATA CCAGACTAAT TTATTTCTAA
ACAAAGTGGC GAATGGTGTC CAATGCCTTT TCTGAGAAAG AGAAAAATGG      159600
CCTCTTCAAA AGACGCAATT TGGTATAGAC TTGTCTGGCC GTTTTGTAGT
GTTCTGGAGT GGGGAGCATA TACTACCCTC ACAGGTAAAC GAAACCTTAC      159700
CTGTATGTAT TGTCTCACCC ATATCCTGCT ACAGGAGAAA AAAAAAAAAA
AAAAAAAAAG AACATTTCCA TGGATCAGTT AGATTCCTTT AGACTTACAG      159800
GGAGGCTAAA ATCACAGCAT TAGGAAAGAC TTTAGCATTT GGCACCTTGA
ATGAATATGG AAAATATAAG ACCCAGGGTC TGTTGGTAGC CAGAAGAATG      159900
GATGCATTTC AACAGCAGTA ATTTGCATGC GTGTATCACT TTGTCACTTT
GAAAGCATTG TGACTCACAT TTCACACTGG ATGTTCACAA AGACTCTGAT      160000
TTAAGCAGGG GTTGTTTTAC AAACGAGAAA ATTGGGATTC AAAGACACAG
GTGGGTCTTG GACGGCAATC TGCTGATTTT TTTTTTCTAG GACTGTCTTT      160100
GAAATATCAT CTGGTCTCCT TGTTAAGATT TAAACTGCTG TGTATCATCA
GAACAATAGC CAATTAACTA ATTGTGGAAA CATTTACAAT CATGCTTCCA      160200
TGTTTTCAAA TTTCCTGTCA CCTAGTTATA AAAGTATGTA TCTCTATACA
TATTATTTGT GGTTGTTTAA GAAGGCAGGA ATTTGACTAG GAATTTTGGT      160300
TCGGTTTAAA TAACCTCCTA GATTATTTAT TATTTGAGCA AAGTTTAAAT
ATATATACAA ATAATTTGAA ATAATATTTA TGATAATGTT TCGATAATAT      160400
ATGCTTCATA CACCTAGCTT GCTTGTTACA GATGAACACC TGACAATAAT
GTTTAGTTTT CATATCCCAA ATAGGAGACT ATCTTCCTTT TTCCTAAAGG      160500
TTTTTATGCC TATTCCAAAC TTGCTTTAAA AGACTTTTTT TTTTCCAGAT
     [EXON 2: 160549..
ATGGAAAACG ATCCAGCCCA GAGACACTGA TTTCAGACCT CTTGATGAGA      160600
GAAAGCACAG AAAATGTTCC CAGAACTCGG TATGACAAGG CTTGTGATGG
              ..160629]
GGACATTGTT GCAGAGCTCA AGGTGCCCAG GGGAGGGAAG TCAGAGACAG      160700
GTGCCTGGTG GGAGGCACCA CCAGGCTTCT AGACTAGGGG AGATTTCGGC
AGAAATGAGG ATGAAGAAGC AGGCAGAGGA GACCTCTCTT TTCATGTACT      160800
CATTGTCAAA TCTCTAAACT TCTGTGGAAT AACATTTTTT TAGCTCAGGT
TATTTGGTGC TCCTTTGACG ATTTTAAAAA TAATTGACTT TTAACTTAGA      160900
AATTATATAG AATTTCTAAC ATTGTTGGAG TCAGGGGAGC ATAGAGTGAG
TCCCTGACTC TAAATTGAAA TGACCAACTG TCCATCTCTG GCTGGGAAGA      161000
TCTGTAGATT TTATATATGT GGAGTGTCCA GATGTCCTAG GAAATGTCCC
ATGTGGTTTT TACTCCAACT TTCCCAGTTT TACATGAAGC TGTAGGTAAA      161100
TTAGGTGCTC TGGATTACTT TTTGACATAT TTTAATTCAA CTTGTGTCAT
ATAGTGATTT ATTACATTCA ACGAATGCAC GTTGAATGAC AGTATTCTTA      161200
GGAGTTTATC TGTAGTCTCA GGACTTTCTT GCAGTTTGGG TCTGTCTTGG
AAAGAGGTGC CCACCAATGT CTAAGTGTTA AACATTGAGC TAAGGTTTTC      161300
TGGTGCCTCC TGGACAAGGA GTGCTGCTGA AGAGGAGCCT GGTCCTCTTC
CATCTTCATA ACCTCCTGAC TCTTCAGTCC TTGCTTCCTG TCTCTAACCA      161400
CAGTAAGATC ATAGTTCACA GGGTCCACCC ACGTCTGTGG CAGCTTGGCT
GCCTTAAACC AAGTTCTACA ATGTCAAGAA CCAAGTTCCC CAGAAGAGGT      161500
CACTACAACA CTTGCTCAGT GAGTGTGACT GACGTGGCTC TCCATCAAGT
TCTCCCTAGA GACTCAGCAG GGTGACCCGC TGGGCCACCC ACCTTCCCTT      161600
CTAACCTCCC TCAGTTCTCT TCTAAAATGG CCAGCCAGCA TGAAGGGTGT
```

FIGURE 1D

```
TAGTATAGAG TTGTGAGATT TGTTTTCTCT AAAGGGAACA ACGATTTGAA      161700
CTGAGATAAA ACTCCAGTGC TCCTGTCCCT GGAGGAACCA GCCCAGACTT
AAATAGCAAA TGCAAGGGAA ACAGGCAGAA AATTTGTAAG GGTTATCTTT      161800
TTTTAATTAA ATAATAATAA TAATAATTAT TATTATTATT CAGAGACAAG
GTCTCCTTCT GTCACTCAGG CCAGAGCACA GTGGTGCAAT CAAAGATCAC      161900
TGCAGCCTCG AGCTCCCAGC CTCAAGAGAT CCTCCCTCCT CAGCCTGCCC
AAGTGCTAGG ATTACAGTTG TGAGCCATAA TGCCTGGCCA GGGTTGTCTT      162000
TAACACTAAG CAGTAAATGT CTTTCCATGT TTCACTGACA TGCATATGAT
GTTTGCCACA GTAAAAAGTT TAGGGTAGAC TATATCGACT TACTGTTAGA      162100
AAAATAACAA CGCAGACCTA CTTAGAACCT TGGTTTTTGT CTTTGTTTTG
TGGCCCAAGG TCACACAGTT AGTGCATGAA GTAAAATTAA TTGCAACCTG      162200
ATTGACAGAC TCCATCAGGT GCAGTTAGAA GGAGACAAAA CTTGACGGTC
TTGAAGTTGC CTATGAGTGC TATTTTATTA TAAAGGTCTT ATTATAGTCT      162300
AATAAGTAGC ATCTTACTAC ATAATAAAGC TGGCAGGAAA AGGAAGCATA
TTTAGAGTTG AGGAACAATA ACTTTCCTGT CACAGCTTCA TGTGATTTCT      162400
GTAATTTTGA TTTTCAAAAG ATCGGGGTGT TCTTCGGCTC CTTTGATACA
TACAGAGCCT TTCAAATGCT CTCTCTGATG TTTGCCCAAG ACGCCTGTCA      162500
ACATCGTGGA ATGCTGCTCA GCTGGGGCTG TGCATATTGT GCCTCCGCCT
CACGATCTGA TATTCCACAT GGCTTCTTAT TTAGAATGCC AACAGGAAAC      162600
TTTTCAACAG TTCCCGGTCA TCTTTCACTT CAGATCTAAA TGTCTCACCC
TTGCTCATAC CTCAGGAAGT TGGGCAGCTT TCCTTACATG CTTTGCTTCT      162700
TATGTTTTAC AGGCTTGAAG ACCCTGCAAT GTGGTGATGG GAAATGAGAC
    [EXON 3: 162713..
                ..162737]
TTGCTCTCTG GCCTTTTCCT ATTTTCAGCC CATATTTCAT CGTGTAAAAC      162800
                                   G
GAGAATCCAC CCATCCTACC AATGCATGCA GCCACTGTGC TGAATTCTGC
AATGTTTTCC TTTGTCATCA TTGTATATAT GTGTGTTTAA ATAAAGTATC      162900
ATGCATTCAA AAGTGTATCC TCCTCAATGA AAAATCTATT ACAATAGTGA
                 G                                   C
GGATTATTTT CGTTAAACTT ATTATTAACA ATTAGTCATT ATAGTCTCCC      163000
TCAACTAGCA GATTCAGGGT TCGTTTCGCT AATATTAAAA CAGCGGCTAT
GAAAGATATG AAGATATGAT TATTCATGAT AACCAAGATC AGCTGTGTGG      163100
TGTGTGTGTA TGTGTGTAAA ACATATAGAT GCATTTGAAG ATATTGATCA
AAGGAATGCT AATTTGGGAA ATTTTTCTGA GCTTTCCCCC CATTAACATT      163200
GGAACCCTTT TGCTAACCTT GTTAACCAGT GCTGTCAGCA GAAGAACTTC
CTGACTAACT AGAGCTTTCT GTGGAATGTC CTTTATCCCT CTGTCTAATA      163300
TGGCAGCCAC TAGCCACATG TGGTTTTTGA GCCCCTGAAT GTAGTAAGTG
CAAATAAAGA GATGAGTTTT AAATTTTATT TAATTTTAAT GAATTTTAAC      163400
TTAAATGGCC ATTAAGTCTC TAGTGGCTAC CATATCAGAC CGTTGAGGTC
TATACATCTA TTTTGGTCAC TACTGTTAGA ATTTGAGTGA GAAATACAAT      163500
GAGTGGGATG ATTATCTTTT CATACAACAT GGGGCATGTT AATTAATTTA
GCTGAAATTT ATTAAACTTC CTCCTTCAAG TACCTACTTG CATGGGCTTA      163600
CTTTAGAATG AATGTTATTA TATTTTTACA TGTCATTTCT AAATCAAAGA
AAATGAGTCT CATCCCATAC TGGTACTTCA TACCACACTG TAAAAAATAT      163700
GTATAAATAT GTTTATATGA CATCTTAAAA CTTTACCAAA AAATGCATGA
ATAGCAATAT TCATCATGAC CTAGGGAATC TTGTAGAATA TACAATATTC      163800
TCTGCATATA AAAAATTATA ACTTTTTTTC AGGCAATGGA GTCAGTGTTG
CTGCAAAATG TAATTTCCTT TAATGAAACT AAGATAAAAT TGTTCCGCAA      163900
TTTCCCTAAA ATTTCCTTTA TTTTTATTTC TTTATTCTTG GAACTAAATA
CCAATATGTA AAGCATTTTT AAAATCTAAA AAGTTAAGTA AGTCATAATT      164000
TATTACACAT CAAAGTTATT AACCTAGTAT ATCTGGGAAT TTTCTTTATT
```

FIGURE 1E

AAAATTAGTA ATACACAAAG AAAGTTGACT GAAAAATCCA TCTTTCAGCA 164100
CAAATAAGCT GAGTACTCAT CACCACGCAT ATAGTCATAT ACCACTTCAC
AACATTTTGG TCAATGATGG CCCACACGTA TAGCAGTGGT CCCATAAGAT 164200
TATGATACTG AATGTTTACT GTACCTTTTC TATGTTTAGA TACAAAAATA
CTTACCATTA TGTCACAATT GCCTATAGTA TTCAGTACAG TCACTTGCTG 164300
TATAGGTTTG TAACCTAGGA GCAATAGGCC ATATCATATA GCCTAGGAGT
ATAGAAAGCC ATACCCATCC AGGTTTGTGT AAGTACATTC ATTCTATGAT 164400
GTTTGCCCAA CAACAAGATC ACCTAACAAT GCATTTCTAA GAAAGTACCC
CTATTGATAA TTGATGCATG ACTGTAGAAC TCTTGATTAC ATGGAGAAGA 164500
GCAAACAGGA ATAATGTATA TTACCCCTTC AGAGAGAATC CAGGAATTCA
ATTTTAAACA AAAATTTTAA TCTCTTACCT CATCAAACCT AAATGCCTAG 164600
AAAACTTAGC ATTAAACAAC CGCCTTTTAC TTATACTCAA CCCAGGCCTT
GATCCCTCTG CTTCTTCACC TCATCTTATT GGAAGGAGGA CTTAACCTTC 164700
ACACCGTGTG GGCCCCATGA ACCTGACTCA GTGACTTTCA AGCTCTCTGA
CCTTCAAATT CACTACATGG CTGGGTTCTC ATCCATGAGA TGGGGATAAT 164800
TATTCCTGCC TAAATTTCCA TGTATACCTT TATGGGAATC AAAATTCAGG
TTAAGATAAC GTTACAAGTA ATAGAAAAGA AGTGGGTAAA AATGAAGTTT 164900
TTGATGACTC CTGTGTTGAA TTCTAAATCC CTAATGAACC AGAATTAGCT
TTCAGAAATA AAAAGATAGC AAACTTTTAA AAAACTGCAA AACTGCATTT 165000
CAGGAGATTC TCTATAGTAA TATTTCTCAA ACATTTTCAG GTTTAGAATG
GGAATAGGTT TGAAATACGT TTAAATAGGT TATAGAATAG GTTTAAAGGC 165100
ACTGGATCAC TCAGCTGTCA TTGCAAACAA TAATGATAAT GCTTACTAGA
GTTATAGTAA GAACAATCTA GCCTTCATTA ATGAGATCCC TATGAGCAAC 165200
ATAATCATTA TCAAGTAAAT CCATCAGCAC TTATTAAGGT CTCCTCTCAC
CCTAGCTCTA GTCCATCATC TATGAGGTAT AATGTGTGCA TAAATTTTGC 165300
AGGTGTCATG TTTTTATTTT GATTCCTGCA ATACCCAAGA CATACTCAAG
CAGATGAATA CTCCCCCAAT TTTCACAGAA AGGCCTATAG GAAGTAAGTG 165400
ACTTATCCTA CCTCATTCAA GTTAGTGGGA CAAGACAGCA GCTCCTCTTT
ACTTGAAAAC TTTGCCCTGC TTGACAATGT CTTCCAGCAG TGAATGCAAA 165500
TGGAGATCAG GTTGTGTATG ATCTAGTTAA GAGCATAACT CTGGTGTACC
ACACTGAAAA CAAATTAGAG AGTAACGAAG TACTAAACAA ATTGTGTGAG 165600
AACTATTGAG AAAAGGAAAT TAGTGTGGGC CGACTAGTTG AAAAATGCTT
ACCAGAGAAT GTGGATTGAC CTGAGAATAG GAGGTTTTGG GTCTGGGCTA 165700
AAAACTGTGT GGATTAGCAC AGCTTGTGAA GGGAACCTTC AGAGCTCTAC
TTGACAGGAA CAGAGAGAGA GTAGCAGAGG CTAGATTGAT GAACTAGGTC 165800
TAAGAATGGC CAAACTCATT CATTCACTTA ACATATTTTC ATCACCTAAT
ATGTATTGGT ACTAAGCACT GAGGATACAA TTGTGCTAAG CAGTGAGGAT 165900
ACAAAGATTA GTTTCTCAGG GAATTTACTA TCTAATGAGG GAGGCAAATA
ATTACAGTAC ATTGTTGGAA GTGCAGAGAT GTGTCAGAGG TCTAAGAAGT 166000
AGTTAGGTGC TAGAACCTTC CCACTCATCA TGCTTCCCTT TGCAGTTTCC
TAACAAGGGA GCCGTACATA TAAGAGTCAG TTGCCAACTA GGAAACAGAG 166100
CCAGGTGAAC AAGCATCCAA TGAGGCCCAA ATGTAGGGCT GCCTGGTAAG
CTGAATCAGG GTGCCTGGCT GCCGCCCAAG CCTCCAGAGC ACCCCGTTTC 166200
CAAACCAAAT CTCAGCAGCC TAGTGAAGAG TGACAGGAAA AGTAAGGTTT
CTGGGGAAGA TTCCCAAATT CACTTACACA AATAGCCACA TAGATCAGAA 166300
GGGCCAGCCT GTCACTTCTA GCAGCAGATG GAAAAAGAGA AACGCCACTT
GTGACTGCTC ATTACCTTTC AGTTGTGTCC CTGCCTGTTA AGACAATGTG 166400
GCCTAATGAA TCAGTCAGAA TAGCAACTGT CTCCCATGCA GAGGGAGACA
GGAAAAGAGA AAGAACAGAG GCCAGAGGTC TCATGGGTTA AGGCTAAAAT 166500
GTGGACATCC TATAGACCAT TCAAATTCAA TATGTTGGCT GGGCACGGTG
GCTCACAGCT GTAATCCCAG CACTTTAGGA GGCCAAGGCA GGCGGATCAT 166600
CTGAGATCAG GAGTTTGAGA CCAGCCTGGC CAACATGGTG AAACCCCATC

FIGURE 1F

```
TCCACTAAAA ATACAAAAAT TACCTGGGTA TGGTGGCACA TGCCTGTAGT      166700
CCCAGGTACT TGGGAGGCTG AGACAGGAGA ATCACTT                    166737
```

FIGURE 1G

POLYMORPHISMS IN THE CODING SEQUENCE OF NPY

```
ATGCTAGGTA ACAAGCGACT GGGGCTGTCC GGACTGACCC TCGCCCTGTC
                   C
CCTGCTCGTG TGCCTGGGTG CGCTGGCCGA GGCGTACCCC TCCAAGCCGG      100
                                  A
ACAACCCGGG CGAGGACGCA CCAGCGGAGG ACATGGCCAG ATACTACTCG
GCGCTGCGAC ACTACATCAA CCTCATCACC AGGCAGAGAT ATGGAAAACG      200
ATCCAGCCCA GAGACACTGA TTTCAGACCT CTTGATGAGA GAAAGCACAG
AAAATGTTCC CAGAACTCGG CTTGAAGACC CTGCAATGTG GTGA            294
```

FIGURE 2

ISOFORMS OF THE NPY PROTEIN

MLGNKRLGLS GLTLALSLLV CLGALAEAYP SKPDNPGEDA PAEDMARYYS
P
ALRHYINLIT RQRYGKRSSP ETLISDLLMR ESTENVPRTR LEDPAMW 97

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Nandabalan, Krishnan
Stephens, J. Claiborne
Chew, Anne
Denton, R. Rex
Lanz, Elizabeth M.

<120> DRUG TARGET ISOGENES: POLYMORPHISMS IN THE NEUROPEPTIDE Y GENE

<130> MWH-0035 NPY

<140> tba
<141> 2000-12-21

<150> 60/171,414
<151> 1999-12-21

<160> 89

<170> PatentIn Ver. 2.1

<210> 1
<211> 14537
<212> DNA
<213> Homo sapiens

<400> 1
gagagaagag tggacacgtt tagatacgga agatataaaa aagattcatt gctttatatt 60
gtcctgtaac ttaaaaaaaa gattcaagtt gaaactctag agatgaaagc taccaggtat 120
aaggtgcaaa atacactgga taggattggc agtaatcaga tattacagaa gaaaacatta 180
gtgaactcac agacatagca atagagacta tccaaaataa aatatgaagg aaaaaatgag 240
caaaacagaa gcaaaaacag atcaccaaca aactgtgggc caatatacac tcaccgtctg 300
actcagccat tctggtttta agtattttcc aagagaatta aaggctttat gtccatacaa 360
agacttgcat atgaatgtca aggcagcttt attcgtaaaa gtttcaagtg aaaataacca 420
aaaattcatg agcagtagaa ttgattaaaa attttgaagt atatcatata atgaaatact 480
gcccagcaat aaaaaagaat aaactattgc tacatttaac aatgtgagtg gattcaaaaa 540
taatgatgct aagtgaaaga agtcagacaa agagtaccta ccgtggaact atatttatgt 600
gcaattctag aaaacgcaaa caaatctcta gtgatagcag atcaatggag gagggggtga 660
ggaagaggca aaaggaagga attacaaagg tacatgtgga aatttgggga agtgatgtat 720
aatcttcatt ttcttggttt cagtgatggc ttcctgagca tataacacac atgcatatgt 780
gtaagaatat atccaattga atattttaca catttgtagc tttgtatata ttcagcagta 840
tctattcatt ccttctgctt ctttacctta ggtttagtca tgaaaccact gggtttcatg 900
tggtccaggt tctatctgaa tcaagctatg aacctgctat cagtaagacc tcctaattgc 960
aaaccacatg gacctttacc ggttatttga cttctctcgg tcattacctt ctctagtctc 1020
cactgattca cctccttttg aatctcaaca agctcaacag tatctcaatg ttatcttaac 1080

aaagctgtta gaaataaaca acttgacttt gaaacaaagg agagaaatgg ctcagcagct 1140
aggagaagat gcagggttga gttgccaaca gttggttttg tttgtatttt ggccagggga 1200
tgtggcttgg actggagaga aaggagataa ggatgtaagc acatgtaggg catatcaccc 1260
cctatttttt attctctgaa tccttaaccc tcagaataag ttcttattct tgagaatcaa 1320
tgacattatc ttaagctaaa ttaatcaagc ctccacagtg ttcttctctc aatagtggtg 1380
tgggccttcc tagaagtaat ttttcccaaa ttcagtgata cattttaagt tcagatttta 1440
attgatatga atctgtgata cactctaaaa taagattatt ttattgaaaa gtggactgta 1500
actttccctt tatctaggaa gagctctaag ttagaagatg ttttgcactt ttaccgaagg 1560
ctgtgtcttg taagcacccc cgagcaactc tgagagcctt gatttttgtg tcctcagcat 1620
atgtttgtgt aatacagaaa gagaagcagt tgccaagtga aagggatgtt ggtctccaaa 1680
attatagttt gatcccacaa acacacaaac acatacatgc aaaggattgt ttgcttcacg 1740
gtttttgata tttaattcaa tgctgttgga acagcacaaa aactaagtgt cagtttaaca 1800
gaatcacttg tccttttagc attaaaataa catggaactt aatgctttaa tttcccaaca 1860
tgccttttta tttagaaaga ttcagactta tatttcattt agaaataaaa tgccatttta 1920
tttagaaaga tacaggagca ttcattcacg gaactttcag atctcagtcc actgcataaa 1980
atcttgatcc tgtaataata gtttctgtat cttgcatatt cattcaacag gtttaacgcg 2040
atgagcaaat taatgttcat cgtttttaac atgtttcatc ttaatcagaa cccacattct 2100
caacgttaat tgaacgtaca taggactata caagggttag taaataagac agaaactgtt 2160
gttcatttaa ccaccgtcac tttggaccaa aaaagaaaaa atatatattt ttaaaattga 2220
gcttaaaaga gtctctagaa gctggaagcg tggctctttt tcagcaaact gggggaatag 2280
gtttaccgtg ttccccctct ggggaatttt gagtcgccac actcatgtct cgaccgagcc 2340
tggctcgctg cgtctgagcg agtacttgag gaaggctgat ctagaaaaac cagctgagag 2400
aaggggcaga agcccctgaa accacgggcg ggggtggggt ggggagcgca gctttgggac 2460
cctctagccg gagacttccg gcagctgcct ccgacttgtt ctaagtacag gaaaaatctg 2520
tgcgcccagt tgcctcactc caacagcgcg cagttgtgcc cggcgaggat gccgcgctag 2580
tcgtggagat gccccaccac aaagaggatt caggtgcttc ctactccggc acccagtggg 2640
ttggtagtcc tgttggcagg agacaagaat cgtctgggct gctcctatct ctggcaggac 2700
tagacggggc gtgaaggaaa gaaggaaaga aggaaagcag ggatcgggca ctgcccgagg 2760
gcagatactt gggctttggt gttgtccagc gcgctcggag tgcgctgcct cgctcacgcg 2820
gtcccaggcc ccgcttcttc aggcagtgcc tggggcggga gggttggggt gtgggtggct 2880
ccctaagtcg acactcgtgc ggctgcggtt ccagcccccc ccccccgcca ctcaggggcg 2940
ggaagtggcg ggtgggagtc acccaagcgt gactgcccga ggcccctcct gccgcggcga 3000
ggaagctcca taaaagccct gtcgcgaccc gctctctgca ccccatccgc tggctctcac 3060
ccctcggaga cgctcgcccg acagcatagt acttgccgcc cagccacgcc cgcgcgccag 3120
ccaccgtgag tgctacgacc cgtctgtcta ggggtgggag cgaacggggc gcccgcgaac 3180
ttgctagaga cgcagcctcc cgctctgtgg agccctgggg ccctgggatg atcgcgctcc 3240
actccccagc ggactatgcc ggctccgcgc cccgacgcgg accagccctc ttggcggcta 3300
aattccactt gttcctctgc tcccctctga ttgtccacgg cccttctccc gggcccttcc 3360
cgctgggcgg ttcttctgag ttacctttta gcagatatgg agggagaacc cgggaccgct 3420
atcccaaggc agctggcggt ctccctgcgg gtcgccgcct tgaggcccag gaagcggtgc 3480
gcggtaggaa ggtttccccg gcagcgccat cgagtgagga atccctggag ctctagagcc 3540
ccgcgccctg ccacctccct ggattcttgg gctccaaatc tctttggagc aattctggcc 3600
cagggagcaa ttctctttcc ccttccccac cgcagtcgtc accccgaggt gatctctgct 3660
gtcagcgttg atcccctgaa gctaggcaga ccagaagtaa cagagaagaa acttttcttc 3720
ccagacaaga gtttgggcaa gaagggagaa aagtgaccca gcaggaagaa cttccaattc 3780
ggttttgaat gctaaactgg cggggccccc accttgcact ctcgccgcgc gcttcttggt 3840
ccctgagact tcgaacgaag ttgcgcgaag ttttcaggtg gagcagaggg gcaggtcccg 3900
accggacggc gcccggagcc cgcaaggtgg tgctagccac tcctgggttc tctctgcggg 3960

actgggacga gagcggattg ggggtcgcgt gtggtagcag gaggaggagc gcgggggca 4020
gaggagggag gtgctgcgcg tgggtgctct gaatccccaa gcccgtccgt tgagccttct 4080
gtgcctgcag atgctaggta acaagcgact ggggctgtcc ggactgaccc tcgccctgtc 4140
cctgctcgtg tgcctgggtg cgctggccga ggcgtacccc tccaagccgg acaacccggg 4200
cgaggacgca ccagcggagg acatggccag atactactcg gcgctgcgac actacatcaa 4260
cctcatcacc aggcagaggt gggtgggacc gcgggaccga ttccgggagc gccagtgcct 4320
gcacaccagg agatcctggg gatgttaggg aaagggattg tttcttttcc ttcgctctat 4380
cccagggcag gacagtatca ggcacttagt cagctctagg taaatgtttg tacagggcac 4440
actctacaca aaatgggtac cttccatttt gtgcaactac agtcacagag tcgtgatccc 4500
cagattcagg ttccccaggc tggtaggctg gcaatctcct ctcactcacc tcttatggtt 4560
tgttgtggtt cttacggcag tggggcccgg tccagaaatc tcgaaagtac ccagtgaaag 4620
gggcaagaat gcgccagaga aatgctgtag ggggaaacgc tagcaaggtg tctaggagaa 4680
acagaacgac caccaaagaa aaccaaacca aggagtaaac tgcagggttg ccacagacat 4740
tgtcagactt tccggcctgc ccagggctaa ttgaatgaca gtatttagaa aaagacaaat 4800
agagctattt tcttttcttc ttcactgtgt ttttttgca gtctcactcc catatttgaa 4860
aattctagtc ccaataatta gagatttagc ctgaaaacag tttatctaaa atgcagaggt 4920
ctccctattt tatagctggt gaattctgaa aagaaagtgc ccacttggta tttcattaag 4980
aagaaactta cgttggtcca gtaactttag gatgaattca gggattacat ggaatctgag 5040
aaattgtaaa ctttattctg aacattccga aagtcaaaat gaaatgccat attttcactc 5100
tcaacattga caattaaggg aaaaatcaaa tccaaactat tttacaaatt acataggatt 5160
taaattttgt tttcaaaatt ggttgatctt agttctaaat catttaatat cttattttaa 5220
ttactgagat taaatgattt tagtcatttg accgtatcta aaggataagc cctttagaag 5280
cattaatgtt agttaacatg tttaaattat atgtttaaag tattgtgggg tttatatatt 5340
tatttattta gtccatattt ttactgttta caagtcaggg agtagttgga aaggatccaa 5400
aaggggggaa gatttatact tttaaaataa agtttgcttt ttcctgtttc atcaaaaagg 5460
tatttaacat cctaaaacac ttttgaaaaa gccccaattt atatagaatc tgaaaaacat 5520
gaaaatgcca gaattaaata gtagttgttt gtgatccctt gaaatctgat cattttaatt 5580
agcaaaaagc aataatattt tatattccat acttcccaga aagttaattg attcaaattc 5640
ttgctatgtt atgccgtcat gcagatcaaa gagaacttgt acaaacaaaa agatattaac 5700
tgtttttttct agaaaaacct ttgttagtta tttagctgta gcataataca cagtagaact 5760
catttttccg aagatgattg ccataaagct ttaccatttt acaaagctct agcattttaa 5820
aatgttttcc taggcattct ctcattcagt cgttacctca aagctttgaa gggagtataa 5880
ttatcttcat ttcacaaatc atgaagctaa tgacatttca tggaagctat ttttatagga 5940
attatggaat tggagccctc ctttgagtgt ctttggatgt tgtcaccacc atcaacaggc 6000
agtgttatct ccaacatgtt tgatgagaaa aaaacataaa ggaaagggac ttgggtttca 6060
tggcaagaga ttcagcttag atgggaatga gctggtctct gaggcagagt ttcctgtctc 6120
atctttgtat aaactccttc ttagaaacca gcagaggctc ttaatgtacg tccatcttct 6180
ccttcccacc acagacaccc actcctatgt gcgtttctga aaattacagg gtcatttgtg 6240
gatgtgcttt aatttagtat agatataatc tgatatcatt atctttcttt ccccccgccc 6300
cccgccacca gcattccccc tgcactcaat ttgtgcttat attcctttg ctgtatatat 6360
acccctggct cattgcttct ggaagctgca ttacttattc tagagtgtgg gtctcctaca 6420
atttatttat atagcccccct aaatcagtgg ttctcaactg ggggtgattt tatgctcccc 6480
aggtgtgatt tggcagtgtc tggagacatt tttggttttc acaactggag gtgggcactg 6540
ctggtaccta ggtgtataga tgccaaggag gctgctaaat atcctaccat gcccaggaca 6600
gcccccatta cacagaatag ccagtcccag agaaactcag ccttagatta gtgtttctcc 6660
ttcctggctg cacattaaaa ttatcagggg agcttttaaa atacactgat gctgggatcc 6720
acccctgacc aattaaatca gaccctctgg gatggggcct gggctttaga gtgtgttttt 6780
aacactcccc aggtgattct aatgtgcagc caggtttgag aatctttact ctaaggactc 6840

aatctttgaa agttacagca ttgtagaggg gaaggatgat ccagaattaa ttaaagaagg 6900
aagaagagtg aacagtactg aactggatgc aagattatga ctgtgcaaat cctcctattt 6960
cagatttagg ttctttactg gtaaattgga taaaaagttg gcagttcttt tactcactgg 7020
agcatgatga agataaatga gttacagttg ccctagactt ttacgattag taaaaccaat 7080
ggatggacag aataaataag tataagcaga cacctatagt ctaatatcgc aactccagaa 7140
aatctttgag aaaatctcac acatgatggg agtgggacag ggagttgttt tgtttttgct 7200
tcccttgagt ctcagcaaaa gaacagtcat agaaaatatt tccaaaaaat ttagatataa 7260
ttttgcttgg caattagaat ataccctctg tcagctttgt tcagttgttg gacctattaa 7320
ccaatccata ccagactaat ttatttctaa acaaagtggc gaatggtgtc caatgccttt 7380
tctgagaaag agaaaaatgg cctcttcaaa agacgcaatt tggtatagac ttgtctggcc 7440
gttttgtagt gttctggagt ggggagcata tactaccctc acaggtaaac gaaaccttac 7500
ctgtatgtat tgtctcaccc atatcctgct acaggagaaa aaaaaaaaaa aaaaaaaaag 7560
aacatttcca tggatcagtt agattccttt agacttacag ggaggctaaa atcacagcat 7620
taggaaagac tttagcattt ggcaccttga atgaatatgg aaaatataag acccagggtc 7680
tgttggtagc cagaagaatg gatgcatttc aacagcagta atttgcatgc gtgtatcact 7740
ttgtcacttt gaaagcattg tgactcacat ttcacactgg atgttcacaa agactctgat 7800
ttaagcaggg gttgttttac aaacgagaaa attgggattc aaagacacag gtgggtcttg 7860
gacggcaatc tgctgatttt ttttttctag gactgtcttt gaaatatcat ctggtctcct 7920
tgttaagatt taaactgctg tgtatcatca gaacaatagc caattaacta attgtggaaa 7980
catttacaat catgcttcca tgttttcaaa tttcctgtca cctagttata aaagtatgta 8040
tctctataca tattatttgt ggttgtttaa gaaggcagga atttgactag gaattttggt 8100
tcggtttaaa taacctccta gattatttat tatttgagca aagtttaaat atatatacaa 8160
ataatttgaa ataatattta tgataatgtt tcgataatat atgcttcata cacctagctt 8220
gcttgttaca gatgaacacc tgacaataat gtttagtttt catatcccaa ataggagact 8280
atcttccttt ttcctaaagg tttttatgcc tattccaaac ttgctttaaa agacttttt 8340
ttttccagat atggaaaacg atccagccca gagacactga tttcagacct cttgatgaga 8400
gaaagcacag aaaatgttcc cagaactcgg tatgacaagg cttgtgatgg ggacattgtt 8460
gcagagctca aggtgcccag gggagggaag tcagagacag gtgcctggtg ggaggcacca 8520
ccaggcttct agactagggg agatttcggc agaaatgagg atgaagaagc aggcagagga 8580
gacctctctt ttcatgtact cattgtcaaa tctctaaact tctgtggaat aacatttttt 8640
tagctcaggt tatttggtgc tcctttgacg attttaaaaa taattgactt ttaacttaga 8700
aattatatag aatttctaac attgttggag tcaggggagc atagagtgag tccctgactc 8760
taaattgaaa tgaccaactg tccatctctg gctgggaaga tctgtagatt ttatatatgt 8820
ggagtgtcca gatgtcctag gaaatgtccc atgtggtttt tactccaact ttcccagttt 8880
tacatgaagc tgtaggtaaa ttaggtgctc tggattactt tttgacatat tttaattcaa 8940
cttgtgtcat atagtgattt attacattca acgaatgcac gttgaatgac agtattctta 9000
ggagtttatc tgtagtctca ggactttctt gcagtttggg tctgtcttgg aaagaggtgc 9060
ccaccaatgt ctaagtgtta aacattgagc taaggttttc tggtgcctcc tggacaagga 9120
gtgctgctga agaggagcct ggtcctcttc catcttcata acctcctgac tcttcagtcc 9180
ttgcttcctg tctctaacca cagtaagatc atagttcaca gggtccaccc acgtctgtgg 9240
cagcttggct gccttaaacc aagttctaca atgtcaagaa ccaagttccc cagaagaggt 9300
cactacaaca cttgctcagt gagtgtgact gacgtggctc tccatcaagt tctccctaga 9360
gactcagcag ggtgacccgc tgggccaccc accttccctt ctaacctccc tcagttctct 9420
tctaaaatgg ccagccagca tgaagggtgt tagtatagag ttgtgagatt tgttttctct 9480
aaagggaaca acgatttgaa ctgagataaa actccagtgc tcctgtccct ggaggaacca 9540
gcccagactt aaatagcaaa tgcaagggaa acaggcagaa aatttgtaag ggttatcttt 9600
ttttaattaa ataataataa taataattat tattattatt cagagacaag gtctccttct 9660
gtcactcagg ccagagcaca gtggtgcaat caaagatcac tgcagcctcg agctcccagc 9720

ctcaagagat cctccctcct cagcctgccc aagtgctagg attacagttg tgagccataa 9780
tgcctggcca gggttgtctt taacactaag cagtaaatgt ctttccatgt ttcactgaca 9840
tgcatatgat gtttgccaca gtaaaaagtt tagggtagac tatatcgact tactgttaga 9900
aaaataacaa cgcagaccta cttagaacct tggtttttgt ctttgttttg tggcccaagg 9960
tcacacagtt agtgcatgaa gtaaaattaa ttgcaacctg attgacagac tccatcaggt 10020
gcagttagaa ggagacaaaa cttgacggtc ttgaagttgc ctatgagtgc tattttatta 10080
taaaggtctt attatagtct aataagtagc atcttactac ataataaagc tggcaggaaa 10140
aggaagcata tttagagttg aggaacaata actttcctgt cacagcttca tgtgatttct 10200
gtaattttga ttttcaaaag atcggggtgt tcttcggctc ctttgataca tacagagcct 10260
ttcaaatgct ctctctgatg tttgcccaag acgcctgtca acatcgtgga atgctgctca 10320
gctggggctg tgcatattgt gcctccgcct cacgatctga tattccacat ggcttcttat 10380
ttagaatgcc aacaggaaac ttttcaacag ttcccggtca tctttcactt cagatctaaa 10440
tgtctcaccc ttgctcatac ctcaggaagt tgggcagctt tccttacatg ctttgcttct 10500
tatgttttac aggcttgaag accctgcaat gtggtgatgg gaaatgagac ttgctctctg 10560
gccttttcct attttcagcc catatttcat cgtgtaaaac gagaatccac ccatcctacc 10620
aatgcatgca gccactgtgc tgaattctgc aatgttttcc tttgtcatca ttgtatatat 10680
gtgtgtttaa ataaagtatc atgcattcaa aagtgtatcc tcctcaatga aaaatctatt 10740
acaatagtga ggattatttt cgttaaactt attattaaca attagtcatt atagtctccc 10800
tcaactagca gattcagggt tcgtttcgct aatattaaaa cagcggctat gaaagatatg 10860
aagatatgat tattcatgat aaccaagatc agctgtgtgg tgtgtgtgta tgtgtgtaaa 10920
acatatagat gcatttgaag atattgatca aaggaatgct aatttgggaa atttttctga 10980
gctttccccc cattaacatt ggaacccttt tgctaacctt gttaaccagt gctgtcagca 11040
gaagaacttc ctgactaact agagctttct gtggaatgtc ctttatccct ctgtctaata 11100
tggcagccac tagccacatg tggtttttga gcccctgaat gtagtaagtg caaataaaga 11160
gatgagtttt aaatttttatt taattttaat gaattttaac ttaaatggcc attaagtctc 11220
tagtggctac catatcagac cgttgaggtc tatacatcta ttttggtcac tactgttaga 11280
atttgagtga gaaatacaat gagtgggatg attatctttt catacaacat ggggcatgtt 11340
aattaattta gctgaaattt attaaacttc ctccttcaag tacctacttg catgggctta 11400
ctttagaatg aatgttatta tatttttaca tgtcatttct aaatcaaaga aaatgagtct 11460
catcccatac tggtacttca taccacactg taaaaaatat gtataaatat gtttatatga 11520
catcttaaaa ctttaccaaa aaatgcatga atagcaatat tcatcatgac ctagggaatc 11580
ttgtagaata tacaatattc tctgcatata aaaaattata actttttttc aggcaatgga 11640
gtcagtgttg ctgcaaaatg taatttcctt taatgaaact aagataaaat tgttccgcaa 11700
tttccctaaa atttccttta tttttatttc tttattcttg gaactaaata ccaatatgta 11760
aagcattttt aaaatctaaa aagttaagta agtcataatt tattacacat caaagttatt 11820
aacctagtat atctgggaat tttctttatt aaaattagta atacacaaag aaagttgact 11880
gaaaaatcca tctttcagca caaataagct gagtactcat caccacgcat atagtcatat 11940
accacttcac aacattttgg tcaatgatgg cccacacgta tagcagtggt cccataagat 12000
tatgatactg aatgtttact gtacctttc tatgtttaga tacaaaaata cttaccatta 12060
tgtcacaatt gcctatagta ttcagtacag tcacttgctg tataggtttg taacctagga 12120
gcaataggcc atatcatata gcctaggagt atagaaagcc atacccatcc aggtttgtgt 12180
aagtacattc attctatgat gtttgcccaa caacaagatc acctaacaat gcatttctaa 12240
gaaagtaccc ctattgataa ttgatgcatg actgtagaac tcttgattac atggagaaga 12300
gcaaacagga ataatgtata ttacccccttc agagagaatc caggaattca attttaaaca 12360
aaaattttaa tctcttacct catcaaacct aaatgcctag aaaacttagc attaaacaac 12420
cgccttttac ttatactcaa cccaggcctt gatccctctg cttcttcacc tcatcttatt 12480
ggaaggagga cttaaccttc acaccgtgtg ggccccatga acctgactca gtgactttca 12540
agctctctga ccttcaaatt cactacatgg ctgggttctc atccatgaga tggggataat 12600

tattcctgcc taaatttcca tgtatacctt tatgggaatc aaaattcagg ttaagataac 12660
gttacaagta atagaaaaga agtgggtaaa aatgaagttt ttgatgactc ctgtgttgaa 12720
ttctaaatcc ctaatgaacc agaattagct ttcagaaata aaaagatagc aaacttttaa 12780
aaaactgcaa aactgcattt caggagattc tctatagtaa tatttctcaa acattttcag 12840
gtttagaatg ggaataggtt tgaaatacgt ttaaataggt tatagaatag gtttaaaggc 12900
actggatcac tcagctgtca ttgcaaacaa taatgataat gcttactaga gttatagtaa 12960
gaacaatcta gccttcatta atgagatccc tatgagcaac ataatcatta tcaagtaaat 13020
ccatcagcac ttattaaggt ctcctctcac cctagctcta gtccatcatc tatgaggtat 13080
aatgtgtgca taaattttgc aggtgtcatg tttttatttt gattcctgca atacccaaga 13140
catactcaag cagatgaata ctcccccaat tttcacagaa aggcctatag gaagtaagtg 13200
acttatccta cctcattcaa gttagtggga caagacagca gctcctcttt acttgaaaac 13260
tttgccctgc ttgacaatgt cttccagcag tgaatgcaaa tggagatcag gttgtgtatg 13320
atctagttaa gagcataact ctggtgtacc acactgaaaa caaattagag agtaacgaag 13380
tactaaacaa attgtgtgag aactattgag aaaaggaaat tagtgtgggc cgactagttg 13440
aaaaatgctt accagagaat gtggattgac ctgagaatag gaggttttgg gtctgggcta 13500
aaaactgtgt ggattagcac agcttgtgaa gggaaccttc agagctctac ttgacaggaa 13560
cagagagaga gtagcagagg ctagattgat gaactaggtc taagaatggc caaactcatt 13620
cattcactta acatattttc atcacctaat atgtattggt actaagcact gaggatacaa 13680
ttgtgctaag cagtgaggat acaaagatta gtttctcagg gaatttacta tctaatgagg 13740
gaggcaaata attacagtac attgttggaa gtgcagagat gtgtcagagg tctaagaagt 13800
agttaggtgc tagaaccttc ccactcatca tgcttccctt tgcagtttcc taacaaggga 13860
gccgtacata taagagtcag ttgccaacta ggaaacagag ccaggtgaac aagcatccaa 13920
tgaggcccaa atgtagggct gcctggtaag ctgaatcagg gtgcctggct gccgcccaag 13980
cctccagagc accccgtttc caaaccaaat ctcagcagcc tagtgaagag tgacaggaaa 14040
agtaaggttt ctggggaaga ttcccaaatt cacttacaca aatagccaca tagatcagaa 14100
gggccagcct gtcacttcta gcagcagatg gaaaaagaga aacgccactt gtgactgctc 14160
attacctttc agttgtgtcc ctgcctgtta agacaatgtg gcctaatgaa tcagtcagaa 14220
tagcaactgt ctcccatgca gagggagaca ggaaagagaa aagaacagag gccagaggtc 14280
tcatgggtta aggctaaaat gtggacatcc tatagaccat tcaaattcaa tatgttggct 14340
gggcacggtg gctcacagct gtaatcccag cactttagga ggccaaggca ggcggatcat 14400
ctgagatcag gagtttgaga ccagcctggc caacatggtg aaaccccatc tccactaaaa 14460
atacaaaaat tacctgggta tggtggcaca tgcctgtagt cccaggtact tgggaggctg 14520
agacaggaga atcactt 14537

<210> 2
<211> 294
<212> DNA
<213> Homo sapiens.

<400> 2
atgctaggta acaagcgact ggggctgtcc ggactcaccc tggccctgtc cctgctcgtg 60
tgcctgggtg cgctggccga ggcgtacccc tccaagccgg acaacccggg cgaggacgca 120
ccagcggagg acatggccag atactactcg gcgctgcgac actacatcaa cctcatcacc 180
aggcagagat atggaaaacg atccagccca gagacactga tttcagacct cttgatgaga 240
gaaagcacag aaaatgttcc cagaactcgg cttgaagacc ctgcaatgtg gtga 294

<210> 3
<211> 97
<212> PRT
<213> Homo sapiens

<400> 3
Met Leu Gly Asn Lys Arg Leu Gly Leu Ser Gly Leu Thr Leu Ala Leu
1 5 10 15

Ser Leu Leu Val Cys Leu Gly Ala Leu Ala Glu Ala Tyr Pro Ser Lys
20 25 30

Pro Asp Asn Pro Gly Glu Asp Ala Pro Ala Glu Asp Met Ala Arg Tyr
35 40 45

Tyr Ser Ala Leu Arg His Tyr Ile Asn Leu Ile Thr Arg Gln Arg Tyr
50 55 60

Gly Lys Arg Ser Ser Pro Glu Thr Leu Ile Ser Asp Leu Leu Met Arg
65 70 75 80

Glu Ser Thr Glu Asn Val Pro Arg Thr Arg Leu Glu Asp Pro Ala Met
85 90 95

Trp

<210> 4
<211> 15
<212> DNA
<213> Homo sapiens

<400> 4
tattcattca acagg 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
tattcatgca acagg 15

<210> 6
<211> 15

<212> DNA
<213> Homo sapiens

<400> 6
cagtgggttg gtagt 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
cagtgggctg gtagt 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
cgggagggtt ggggt 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
cgggaggttt ggggt 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
cagccccctc ccccc 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
cagccccatc ccccc 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
ctgtggagcc ctggg 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapiens

<400> 13
ctgtggatcc ctggg 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapiens

<400> 14
agaaaagtga cccag 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
agaaaagcga cccag 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

<400> 16
cgcccggagc ccgca 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
cgcccggggc ccgca 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
aatccccaag cccgt 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19
aatcccccag cccgt 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
cttgcatatt cattc 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
gttaaacctg ttgaa 15

<210> 22
<211> 15

<212> DNA
<213> Homo sapiens

<400> 22
cttgcatatt catgc 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapiens

<400> 23
gttaaacctg ttgca 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
ggcacccagt gggtt 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
aacaggacta ccaac 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
ggcacccagt gggct 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
aacaggacta ccagc 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
ctggggcggg agggt 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
acccacaccc caacc 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapiens

<400> 30
ctggggcggg aggtt 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
acccacaccc caaac 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapiens

<400> 32
cggttccagc cccct 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
agtggcgggg ggagg 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapiens

<400> 34
cggttccagc cccat 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapiens

<400> 35
agtggcgggg ggatg 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapiens

<400> 36
cccgctctgt ggagc 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapiens

<400> 37
cagggcccca gggct 15

<210> 38
<211> 15

<212> DNA
<213> Homo sapiens

<400> 38
cccgctctgt ggatc 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapiens

<400> 39
cagggcccca gggat 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapiens

<400> 40
gaagggagaa aagtg 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapiens

<400> 41
ttcctgctgg gtcac 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapiens

<400> 42
gaagggagaa aagcg 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
ttcctgctgg gtcgc 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapiens

<400> 44
ggacggcgcc cggag 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapiens

<400> 45
ccaccttgcg ggctc 15

<210> 46
<211> 15
<212> DNA
<213> Homo sapiens

<400> 46
ggacggcgcc cgggg 15

<210> 47
<211> 15
<212> DNA
<213> Homo sapiens

<400> 47
ccaccttgcg ggccc 15

<210> 48
<211> 15
<212> DNA
<213> Homo sapiens

<400> 48
gctctgaatc cccaa 15

<210> 49
<211> 15
<212> DNA
<213> Homo sapiens

<400> 49
caacggacgg gcttg 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapiens

<400> 50
gctctgaatc cccca 15

<210> 51
<211> 15
<212> DNA
<213> Homo sapiens

<400> 51
caacggacgg gctgg 15

<210> 52
<211> 10
<212> DNA
<213> Homo sapiens

<400> 52
gcatattcat 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapiens

<400> 53
aaacctgttg 10

<210> 54
<211> 10

<212> DNA
<213> Homo sapiens

<400> 54
acccagtggg 10

<210> 55
<211> 10
<212> DNA
<213> Homo sapiens

<400> 55
aggactacca 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapiens

<400> 56
gggcgggagg 10

<210> 57
<211> 10
<212> DNA
<213> Homo sapiens

<400> 57
cacaccccaa 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapiens

<400> 58
ttccagcccc 10

<210> 59
<211> 10
<212> DNA
<213> Homo sapiens

<400> 59
ggcgggggga 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapiens

<400> 60
gctctgtgga 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapiens

<400> 61
ggccccaggg 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapiens

<400> 62
gggagaaaag 10

<210> 63
<211> 10
<212> DNA
<213> Homo sapiens

<400> 63
ctgctgggtc 10

<210> 64
<211> 10
<212> DNA
<213> Homo sapiens

<400> 64
cggcgcccgg 10

<210> 65
<211> 10
<212> DNA
<213> Homo sapiens

<400> 65
ccttgcgggc 10

<210> 66
<211> 10
<212> DNA
<213> Homo sapiens

<400> 66
ctgaatcccc 10

<210> 67
<211> 10
<212> DNA
<213> Homo sapiens

<400> 67
cggacgggct 10

<210> 68
<211> 24
<212> DNA
<213> Homo sapiens

<400> 68
ggagcattca ttcacggaac tttc 24

<210> 69
<211> 21
<212> DNA
<213> Homo sapiens

<400> 69
ggctcggtcg agacatgagt g 21

<210> 70
<211> 24

<212> DNA
<213> Homo sapiens

<400> 70
catttaacca ccgtcacttt ggac 24

<210> 71
<211> 23
<212> DNA
<213> Homo sapiens

<400> 71
agtcctgcca gagataggag cag 23

<210> 72
<211> 22
<212> DNA
<213> Homo sapiens

<400> 72
ccagttgcct cactccaaca gc 22

<210> 73
<211> 20
<212> DNA
<213> Homo sapiens

<400> 73
tgactcccac ccgccacttc 20

<210> 74
<211> 21
<212> DNA
<213> Homo sapiens

<400> 74
acttgggctt tggtgttgtc c 21

<210> 75
<211> 22
<212> DNA
<213> Homo sapiens

<400> 75
aatcagaggg gagcagagga ac 22

<210> 76
<211> 23
<212> DNA
<213> Homo sapiens

<400> 76
gaacttgcta gagacgcagc ctc 23

<210> 77
<211> 21
<212> DNA
<213> Homo sapiens

<400> 77
cttcagggga tcaacgctga c 21

<210> 78
<211> 21
<212> DNA
<213> Homo sapiens

<400> 78
tggattcttg ggctccaaat c 21

<210> 79
<211> 20
<212> DNA
<213> Homo sapiens

<400> 79
gcttggggat tcagagcacc 20

<210> 80
<211> 23
<212> DNA
<213> Homo sapiens

<400> 80
gcttcttggt ccctgagact tcg 23

<210> 81
<211> 21
<212> DNA
<213> Homo sapiens

<400> 81
ccctttccct aacatcccca g 21

<210> 82
<211> 26
<212> DNA
<213> Homo sapiens

<400> 82
gcttgcttgt tacagatgaa cacctg 26

<210> 83
<211> 23
<212> DNA
<213> Homo sapiens

<400> 83
cgtcaaagga gcaccaaata acc 23

<210> 84
<211> 24
<212> DNA
<213> Homo sapiens

<400> 84
caggaaactt ttcaacagtt cccg 24

<210> 85
<211> 21
<212> DNA
<213> Homo sapiens

<400> 85
gcgaaacgaa ccctgaatct g 21

<210> 86
<211> 21

<212> DNA
<213> Homo sapiens

<400> 86
gatctcctgg tgtgcaggca c 21

<210> 87
<211> 24
<212> DNA
<213> Homo sapiens

<400> 87
aaatgtctca cccttgctca tacc 24

<210> 88
<211> 24
<212> DNA
<213> Homo sapiens

<400> 88
cgaaccctga atctgctagt tgag 24

<210> 89
<211> 14537
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (2024)
<223> PS1: polymorphic base T or G

<220>
<221> allele
<222> (2641)
<223> PS2: polymorphic base T or C

<220>
<221> allele
<222> (2863)
<223> PS3: polymorphic base G or T

<220>
<221> allele
<222> (2919)

<223> PS4: polymorphic base C or A

<220>
<221> allele
<222> (3212)
<223> PS5: polymorphic base G or T

<220>
<221> allele
<222> (3213)
<223> PS6: polymorphic base C or A

<220>
<221> allele
<222> (3220)
<223> PS7: polymorphic base G or T

<220>
<221> allele
<222> (3362)
<223> PS8: polymorphic base G or A

<220>
<221> allele
<222> (3754)
<223> PS9: polymorphic base T or C

<220>
<221> allele
<222> (3819)
<223> PS10: polymorphic base C or T

<220>
<221> allele
<222> (3839)
<223> PS11: polymorphic base G or A

<220>
<221> allele
<222> (3901)
<223> PS12: polymorphic base A or G

<220>
<221> allele
<222> (3917)
<223> PS13: polymorphic base A or G

<220>

<221> allele
<222> (4059)
<223> PS14: polymorphic base A or C

<220>
<221> allele
<222> (4110)
<223> PS15: polymorphic base T or C

<220>
<221> allele
<222> (4174)
<223> PS16: polymorphic base G or A

<220>
<221> allele
<222> (10582)
<223> PS17: polymorphic base A or G

<220>
<221> allele
<222> (10717)
<223> PS18: polymorphic base A or G

<220>
<221> allele
<222> (10748)
<223> PS19: polymorphic base T or C

<400> 89
gagagaagag tggacacgtt tagatacgga agatataaaa aagattcatt gctttatatt 60
gtcctgtaac ttaaaaaaaa gattcaagtt gaaactctag agatgaaagc taccaggtat 120
aaggtgcaaa atacactgga taggattggc agtaatcaga tattacagaa gaaaacatta 180
gtgaactcac agacatagca atagagacta tccaaaataa aatatgaagg aaaaaatgag 240
caaaacagaa gcaaaaacag atcaccaaca aactgtgggc caatatacac tcaccgtctg 300
actcagccat tctggtttta agtattttcc aagagaatta aaggctttat gtccatacaa 360
agacttgcat atgaatgtca aggcagcttt attcgtaaaa gtttcaagtg aaaataacca 420
aaaattcatg agcagtagaa ttgattaaaa attttgaagt atatcatata atgaaatact 480
gcccagcaat aaaaaagaat aaactattgc tacatttaac aatgtgagtg gattcaaaaa 540
taatgatgct aagtgaaaga agtcagacaa agagtaccta ccgtggaact atatttatgt 600
gcaattctag aaaacgcaaa caaatctcta gtgatagcag atcaatggag gaggggtga 660
ggaagaggca aaaggaagga attacaaagg tacatgtgga aatttgggga agtgatgtat 720
aatcttcatt ttcttggttt cagtgatggc ttcctgagca tataacacac atgcatatgt 780
gtaagaatat atccaattga atattttaca catttgtagc tttgtatata ttcagcagta 840
tctattcatt ccttctgctt ctttacctta ggtttagtca tgaaaccact gggtttcatg 900
tggtccaggt tctatctgaa tcaagctatg aacctgctat cagtaagacc tcctaattgc 960
aaaccacatg gacctttacc ggttatttga cttctctcgg tcattacctt ctctagtctc 1020
cactgattca cctccttttg aatctcaaca agctcaacag tatctcaatg ttatcttaac 1080

aaagctgtta gaaataaaca acttgacttt gaaacaaagg agagaaatgg ctcagcagct 1140
aggagaagat gcagggttga gttgccaaca gttggttttg tttgtatttt ggccagggga 1200
tgtggcttgg actggagaga aaggagataa ggatgtaagc acatgtaggg catatcaccc 1260
cctatttttt attctctgaa tccttaaccc tcagaataag ttcttattct tgagaatcaa 1320
tgacattatc ttaagctaaa ttaatcaagc ctccacagtg ttcttctctc aatagtggtg 1380
tgggccttcc tagaagtaat ttttcccaaa ttcagtgata cattttaagt tcagatttta 1440
attgatatga atctgtgata cactctaaaa taagattatt ttattgaaaa gtggactgta 1500
actttccctt tatctaggaa gagctctaag ttagaagatg ttttgcactt ttaccgaagg 1560
ctgtgtcttg taagcacccc cgagcaactc tgagagcctt gatttttgtg tcctcagcat 1620
atgtttgtgt aatacagaaa gagaagcagt tgccaagtga aagggatgtt ggtctccaaa 1680
attatagttt gatcccacaa acacacaaac acatacatgc aaaggattgt ttgcttcacg 1740
gtttttgata tttaattcaa tgctgttgga acagcacaaa aactaagtgt cagtttaaca 1800
gaatcacttg tccttttagc attaaaataa catggaactt aatgctttaa tttcccaaca 1860
tgccttttta tttagaaaga ttcagactta tatttcattt agaaataaaa tgccatttta 1920
tttagaaaga tacaggagca ttcattcacg gaactttcag atctcagtcc actgcataaa 1980
atcttgatcc tgtaataata gtttctgtat cttgcatatt catkcaacag gtttaacgcg 2040
atgagcaaat taatgttcat cgtttttaac atgtttcatc ttaatcagaa cccacattct 2100
caacgttaat tgaacgtaca taggactata caagggttag taaataagac agaaactgtt 2160
gttcatttaa ccaccgtcac tttggaccaa aaaagaaaaa atatatattt ttaaaattga 2220
gcttaaaaga gtctctagaa gctggaagcg tggctctttt tcagcaaact gggggaatag 2280
gtttaccgtg ttccccctct ggggaatttt gagtcgccac actcatgtct cgaccgagcc 2340
tggctcgctg cgtctgagcg agtacttgag gaaggctgat ctagaaaaac cagctgagag 2400
aaggggcaga agcccctgaa accacgggcg ggggtggggt ggggagcgca gctttgggac 2460
cctctagccg gagacttccg gcagctgcct ccgacttgtt ctaagtacag gaaaaatctg 2520
tgcgcccagt tgcctcactc caacagcgcg cagttgtgcc cggcgaggat gccgcgctag 2580
tcgtggagat gccccaccac aaagaggatt caggtgcttc ctactccggc acccagtggg 2640
ytggtagtcc tgttggcagg agacaagaat cgtctgggct gctcctatct ctggcaggac 2700
tagacggggc gtgaaggaaa gaaggaaaga aggaaagcag ggatcgggca ctgcccgagg 2760
gcagatactt gggctttggt gttgtccagc gcgctcggag tgcgctgcct cgctcacgcg 2820
gtcccaggcc ccgcttcttc aggcagtgcc tggggcggga ggkttggggt gtgggtggct 2880
ccctaagtcg acactcgtgc ggctgcggtt ccagccccmt ccccccgcca ctcaggggcg 2940
ggaagtggcg ggtgggagtc acccaagcgt gactgcccga ggcccctcct gccgcggcga 3000
ggaagctcca taaaagccct gtcgcgaccc gctctctgca ccccatccgc tggctctcac 3060
ccctcggaga cgctcgcccg acagcatagt acttgccgcc cagccacgcc cgcgcgccag 3120
ccaccgtgag tgctacgacc cgtctgtcta ggggtgggag cgaacggggc gcccgcgaac 3180
ttgctagaga cgcagcctcc cgctctgtgg akmcctgggk ccctgggatg atcgcgctcc 3240
actccccagc ggactatgcc ggctccgcgc cccgacgcgg accagccctc ttggcggcta 3300
aattccactt gttcctctgc tcccctctga ttgtccacgg cccttctccc gggcccttcc 3360
crctgggcgg ttcttctgag ttacctttta gcagatatgg agggagaacc cgggaccgct 3420
atcccaaggc agctggcggt ctccctgcgg gtcgccgcct tgaggcccag gaagcggtgc 3480
gcggtaggaa ggtttccccg gcagcgccat cgagtgagga atccctggag ctctagagcc 3540
ccgcgccctg ccacctccct ggattcttgg gctccaaatc tctttggagc aattctggcc 3600
cagggagcaa ttctctttcc ccttccccac cgcagtcgtc accccgaggt gatctctgct 3660
gtcagcgttg atcccctgaa gctaggcaga ccagaagtaa cagagaagaa acttttcttc 3720
ccagacaaga gtttgggcaa gaagggagaa aagygaccca gcaggaagaa cttccaattc 3780
ggttttgaat gctaaactgg cggggccccc accttgcayt ctcgccgcgc gcttcttgrt 3840
ccctgagact tcgaacgaag ttgcgcgaag ttttcaggtg gagcagaggg gcaggtcccg 3900
rccggacggc gcccggrgcc cgcaaggtgg tgctagccac tcctgggttc tctctgcggg 3960

actgggacga gagcggattg ggggtcgcgt gtggtagcag gaggaggagc gcgggggca 4020
gaggagggag gtgctgcgcg tgggtgctct gaatccccma gcccgtccgt tgagccttct 4080
gtgcctgcag atgctaggta acaagcgacy ggggctgtcc ggactgaccc tcgccctgtc 4140
cctgctcgtg tgcctgggtg cgctggccga ggcrtacccc tccaagccgg acaacccggg 4200
cgaggacgca ccagcggagg acatggccag atactactcg gcgctgcgac actacatcaa 4260
cctcatcacc aggcagaggt gggtgggacc gcgggaccga ttccgggagc gccagtgcct 4320
gcacaccagg agatcctggg gatgttaggg aaagggattg tttcttttcc ttcgctctat 4380
cccagggcag gacagtatca ggcacttagt cagctctagg taaatgtttg tacagggcac 4440
actctacaca aaatgggtac cttccatttt gtgcaactac agtcacagag tcgtgatccc 4500
cagattcagg ttccccaggc tggtaggctg gcaatctcct ctcactcacc tcttatggtt 4560
tgttgtggtt cttacggcag tggggcccgg tccagaaatc tcgaaagtac ccagtgaaag 4620
gggcaagaat gcgccagaga aatgctgtag ggggaaacgc tagcaaggtg tctaggagaa 4680
acagaacgac caccaaagaa aaccaaacca aggagtaaac tgcagggttg ccacagacat 4740
tgtcagactt tccggcctgc ccagggctaa ttgaatgaca gtatttagaa aaagacaaat 4800
agagctattt tcttttcttc ttcactgtgt ttttttgca gtctcactcc catatttgaa 4860
aattctagtc ccaataatta gagatttagc ctgaaaacag tttatctaaa atgcagaggt 4920
ctccctattt tatagctggt gaattctgaa aagaaagtgc ccacttggta tttcattaag 4980
aagaaactta cgttggtcca gtaactttag gatgaattca gggattacat ggaatctgag 5040
aaattgtaaa ctttattctg aacattccga aagtcaaaat gaaatgccat attttcactc 5100
tcaacattga caattaaggg aaaaatcaaa tccaaactat tttacaaatt acataggatt 5160
taaattttgt tttcaaaatt ggttgatctt agttctaaat catttaatat cttattttaa 5220
ttactgagat taaatgattt tagtcatttg accgtatcta aaggataagc cctttagaag 5280
cattaatgtt agttaacatg tttaaattat atgtttaaag tattgtgggg tttatatatt 5340
tatttattta gtccatattt ttactgttta caagtcaggg agtagttgga aaggatccaa 5400
aaggggggaa gatttatact tttaaaataa agtttgcttt ttcctgtttc atcaaaaagg 5460
tatttaacat cctaaaacac ttttgaaaaa gccccaattt atatagaatc tgaaaaacat 5520
gaaaatgcca gaattaaata gtagttgttt gtgatccctt gaaatctgat cattttaatt 5580
agcaaaaagc aataatattt tatattccat acttcccaga aagttaattg attcaaattc 5640
ttgctatgtt atgccgtcat gcagatcaaa gagaacttgt acaaacaaaa agatattaac 5700
tgtttttttct agaaaaacct ttgttagtta tttagctgta gcataataca cagtagaact 5760
catttttccg aagatgattg ccataaagct ttaccatttt acaaagctct agcattttaa 5820
aatgttttcc taggcattct ctcattcagt cgttacctca aagctttgaa gggagtataa 5880
ttatcttcat ttcacaaatc atgaagctaa tgacatttca tggaagctat ttttatagga 5940
attatggaat tggagccctc ctttgagtgt ctttggatgt tgtcaccacc atcaacaggc 6000
agtgttatct ccaacatgtt tgatgagaaa aaaacataaa ggaaagggac ttgggtttca 6060
tggcaagaga ttcagcttag atgggaatga gctggtctct gaggcagagt ttcctgtctc 6120
atctttgtat aaactccttc ttagaaacca gcagaggctc ttaatgtacg tccatcttct 6180
ccttcccacc acagacaccc actcctatgt gcgtttctga aaattacagg gtcatttgtg 6240
gatgtgcttt aatttagtat agatataatc tgatatcatt atctttcttt ccccccgccc 6300
cccgccacca gcattccccc tgcactcaat ttgtgcttat attccttttg ctgtatatat 6360
acccctggct cattgcttct ggaagctgca ttacttattc tagagtgtgg gtctcctaca 6420
atttatttat atagcccccct aaatcagtgg ttctcaactg ggggtgattt tatgctcccc 6480
aggtgtgatt tggcagtgtc tggagacatt tttggttttc acaactggag gtgggcactg 6540
ctggtaccta ggtgtataga tgccaaggag gctgctaaat atcctaccat gcccaggaca 6600
gccccatta cacagaatag ccagtcccag agaaactcag ccttagatta gtgtttctcc 6660
ttcctggctg cacattaaaa ttatcagggg agcttttaaa atacactgat gctgggatcc 6720
acccctgacc aattaaatca gaccctctgg gatggggcct gggctttaga gtgtgttttt 6780
aacactcccc aggtgattct aatgtgcagc caggtttgag aatctttact ctaaggactc 6840

aatctttgaa agttacagca ttgtagaggg gaaggatgat ccagaattaa ttaaagaagg 6900
aagaagagtg aacagtactg aactggatgc aagattatga ctgtgcaaat cctcctattt 6960
cagatttagg ttctttactg gtaaattgga taaaaagttg gcagttcttt tactcactgg 7020
agcatgatga agataaatga gttacagttg ccctagactt ttacgattag taaaaccaat 7080
ggatggacag aataaataag tataagcaga cacctatagt ctaatatcgc aactccagaa 7140
aatctttgag aaaatctcac acatgatggg agtgggacag ggagttgttt tgtttttgct 7200
tcccttgagt ctcagcaaaa gaacagtcat agaaaatatt tccaaaaaat ttagatataa 7260
ttttgcttgg caattagaat ataccctctg tcagctttgt tcagttgttg gacctattaa 7320
ccaatccata ccagactaat ttatttctaa acaaagtggc gaatggtgtc caatgccttt 7380
tctgagaaag agaaaaatgg cctcttcaaa agacgcaatt tggtatagac ttgtctggcc 7440
gttttgtagt gttctggagt ggggagcata tactaccctc acaggtaaac gaaaccttac 7500
ctgtatgtat tgtctcaccc atatcctgct acaggagaaa aaaaaaaaaa aaaaaaaaag 7560
aacatttcca tggatcagtt agattccttt agacttacag ggaggctaaa atcacagcat 7620
taggaaagac tttagcattt ggcaccttga atgaatatgg aaaatataag acccagggtc 7680
tgttggtagc cagaagaatg gatgcatttc aacagcagta atttgcatgc gtgtatcact 7740
ttgtcacttt gaaagcattg tgactcacat ttcacactgg atgttcacaa agactctgat 7800
ttaagcaggg gttgttttac aaacgagaaa attgggattc aaagacacag gtgggtcttg 7860
gacggcaatc tgctgatttt ttttttctag gactgtcttt gaaatatcat ctggtctcct 7920
tgttaagatt taaactgctg tgtatcatca gaacaatagc caattaacta attgtggaaa 7980
catttacaat catgcttcca tgttttcaaa tttcctgtca cctagttata aaagtatgta 8040
tctctataca tattatttgt ggttgtttaa gaaggcagga atttgactag gaattttggt 8100
tcggtttaaa taacctccta gattatttat tatttgagca aagtttaaat atatatacaa 8160
ataatttgaa ataatattta tgataatgtt tcgataatat atgcttcata cacctagctt 8220
gcttgttaca gatgaacacc tgacaataat gtttagtttt catatcccaa ataggagact 8280
atcttccttt ttcctaaagg tttttatgcc tattccaaac ttgctttaaa agacttttt 8340
ttttccagat atggaaaacg atccagccca gagacactga tttcagacct cttgatgaga 8400
gaaagcacag aaaatgttcc cagaactcgg tatgacaagg cttgtgatgg ggacattgtt 8460
gcagagctca aggtgcccag gggagggaag tcagagacag gtgcctggtg ggaggcacca 8520
ccaggcttct agactagggg agatttcggc agaaatgagg atgaagaagc aggcagagga 8580
gacctctctt ttcatgtact cattgtcaaa tctctaaact tctgtggaat aacatttttt 8640
tagctcaggt tatttggtgc tcctttgacg attttaaaaa taattgactt ttaacttaga 8700
aattatatag aatttctaac attgttggag tcaggggagc atagagtgag tccctgactc 8760
taaattgaaa tgaccaactg tccatctctg gctgggaaga tctgtagatt ttatatatgt 8820
ggagtgtcca gatgtcctag gaaatgtccc atgtggtttt tactccaact ttcccagttt 8880
tacatgaagc tgtaggtaaa ttaggtgctc tggattactt tttgacatat tttaattcaa 8940
cttgtgtcat atagtgattt attacattca acgaatgcac gttgaatgac agtattctta 9000
ggagtttatc tgtagtctca ggactttctt gcagtttggg tctgtcttgg aaagaggtgc 9060
ccaccaatgt ctaagtgtta aacattgagc taaggttttc tggtgcctcc tggacaagga 9120
gtgctgctga agaggagcct ggtcctcttc catcttcata acctcctgac tcttcagtcc 9180
ttgcttcctg tctctaacca cagtaagatc atagttcaca gggtccaccc acgtctgtgg 9240
cagcttggct gccttaaacc aagttctaca atgtcaagaa ccaagttccc cagaagaggt 9300
cactacaaca cttgctcagt gagtgtgact gacgtggctc tccatcaagt tctcctaga 9360
gactcagcag ggtgacccgc tgggccaccc accttccctt ctaacctccc tcagttctct 9420
tctaaaatgg ccagccagca tgaagggtgt tagtatagag ttgtgagatt tgttttctct 9480
aaagggaaca acgatttgaa ctgagataaa actccagtgc tcctgtccct ggaggaacca 9540
gcccagactt aaatagcaaa tgcaagggaa acaggcagaa aatttgtaag ggttatcttt 9600
ttttaattaa ataataataa taataattat tattattatt cagagacaag gtctccttct 9660
gtcactcagg ccagagcaca gtggtgcaat caaagatcac tgcagcctcg agctcccagc 9720

ctcaagagat cctccctcct cagcctgccc aagtgctagg attacagttg tgagccataa 9780
tgcctggcca gggttgtctt taacactaag cagtaaatgt ctttccatgt ttcactgaca 9840
tgcatatgat gtttgccaca gtaaaaagtt tagggtagac tatatcgact tactgttaga 9900
aaaataacaa cgcagaccta cttagaacct tggtttttgt ctttgttttg tggcccaagg 9960
tcacacagtt agtgcatgaa gtaaaattaa ttgcaacctg attgacagac tccatcaggt 10020
gcagttagaa ggagacaaaa cttgacggtc ttgaagttgc ctatgagtgc tattttatta 10080
taaaggtctt attatagtct aataagtagc atcttactac ataataaagc tggcaggaaa 10140
aggaagcata tttagagttg aggaacaata actttcctgt cacagcttca tgtgatttct 10200
gtaattttga ttttcaaaag atcggggtgt tcttcggctc ctttgataca tacagagcct 10260
ttcaaatgct ctctctgatg tttgcccaag acgcctgtca acatcgtgga atgctgctca 10320
gctggggctg tgcatattgt gcctccgcct cacgatctga tattccacat ggcttcttat 10380
ttagaatgcc aacaggaaac ttttcaacag ttcccggtca tctttcactt cagatctaaa 10440
tgtctcaccc ttgctcatac ctcaggaagt tgggcagctt tccttacatg ctttgcttct 10500
tatgttttac aggcttgaag accctgcaat gtggtgatgg gaaatgagac ttgctctctg 10560
gccttttcct attttcagcc crtatttcat cgtgtaaaac gagaatccac ccatcctacc 10620
aatgcatgca gccactgtgc tgaattctgc aatgttttcc tttgtcatca ttgtatatat 10680
gtgtgtttaa ataaagtatc atgcattcaa aagtgtrtcc tcctcaatga aaaatctatt 10740
acaatagyga ggattatttt cgttaaactt attattaaca attagtcatt atagtctccc 10800
tcaactagca gattcagggt tcgtttcgct aatattaaaa cagcggctat gaaagatatg 10860
aagatatgat tattcatgat aaccaagatc agctgtgtgg tgtgtgtgta tgtgtgtaaa 10920
acatatagat gcatttgaag atattgatca aaggaatgct aatttgggaa atttttctga 10980
gctttccccc cattaacatt ggaacccttt tgctaacctt gttaaccagt gctgtcagca 11040
gaagaacttc ctgactaact agagctttct gtggaatgtc ctttatccct ctgtctaata 11100
tggcagccac tagccacatg tggtttttga gcccctgaat gtagtaagtg caaataaaga 11160
gatgagtttt aaattttatt taattttaat gaattttaac ttaaatggcc attaagtctc 11220
tagtggctac catatcagac cgttgaggtc tatacatcta ttttggtcac tactgttaga 11280
atttgagtga gaaatacaat gagtgggatg attatctttt catacaacat ggggcatgtt 11340
aattaattta gctgaaattt attaaacttc ctccttcaag tacctacttg catgggctta 11400
ctttagaatg aatgttatta tatttttaca tgtcatttct aaatcaaaga aaatgagtct 11460
catcccatac tggtacttca taccacactg taaaaaatat gtataaatat gtttatatga 11520
catcttaaaa ctttaccaaa aaatgcatga atagcaatat tcatcatgac ctagggaatc 11580
ttgtagaata tacaatattc tctgcatata aaaaattata actttttttc aggcaatgga 11640
gtcagtgttg ctgcaaaatg taatttcctt taatgaaact aagataaaat tgttccgcaa 11700
tttccctaaa atttccttta tttttatttc tttattcttg gaactaaata ccaatatgta 11760
aagcattttt aaaatctaaa aagttaagta agtcataatt tattacacat caaagttatt 11820
aacctagtat atctgggaat tttctttatt aaaattagta atacacaaag aaagttgact 11880
gaaaaatcca tctttcagca caaataagct gagtactcat caccacgcat atagtcatat 11940
accacttcac aacattttgg tcaatgatgg cccacacgta tagcagtggt cccataagat 12000
tatgatactg aatgtttact gtacctttte tatgtttaga tacaaaaata cttaccatta 12060
tgtcacaatt gcctatagta ttcagtacag tcacttgctg tataggtttg taacctagga 12120
gcaataggcc atatcatata gcctaggagt atagaaagcc atacccatcc aggtttgtgt 12180
aagtacattc attctatgat gtttgcccaa caacaagatc acctaacaat gcatttctaa 12240
gaaagtaccc ctattgataa ttgatgcatg actgtagaac tcttgattac atggagaaga 12300
gcaaacagga ataatgtata ttacccctte agagagaatc caggaattca attttaaaca 12360
aaaattttaa tctcttacct catcaaacct aaatgcctag aaaacttagc attaaacaac 12420
cgccttttac ttatactcaa cccaggcctt gatccctctg cttcttcacc tcatcttatt 12480
ggaaggagga cttaaccttc acaccgtgtg ggccccatga acctgactca gtgactttca 12540
agctctctga ccttcaaatt cactacatgg ctgggttctc atccatgaga tggggataat 12600

tattcctgcc taaatttcca tgtatacctt tatgggaatc aaaattcagg ttaagataac 12660
gttacaagta atagaaaaga agtgggtaaa aatgaagttt ttgatgactc ctgtgttgaa 12720
ttctaaatcc ctaatgaacc agaattagct ttcagaaata aaaagatagc aaacttttaa 12780
aaaactgcaa aactgcattt caggagattc tctatagtaa tatttctcaa acattttcag 12840
gtttagaatg ggaataggtt tgaaatacgt ttaaataggt tatagaatag gtttaaaggc 12900
actggatcac tcagctgtca ttgcaaacaa taatgataat gcttactaga gttatagtaa 12960
gaacaatcta gccttcatta atgagatccc tatgagcaac ataatcatta tcaagtaaat 13020
ccatcagcac ttattaaggt ctcctctcac cctagctcta gtccatcatc tatgaggtat 13080
aatgtgtgca taaattttgc aggtgtcatg tttttatttt gattcctgca atacccaaga 13140
catactcaag cagatgaata ctcccccaat tttcacagaa aggcctatag gaagtaagtg 13200
acttatccta cctcattcaa gttagtggga caagacagca gctcctcttt acttgaaaac 13260
tttgccctgc ttgacaatgt cttccagcag tgaatgcaaa tggagatcag gttgtgtatg 13320
atctagttaa gagcataact ctggtgtacc acactgaaaa caaattagag agtaacgaag 13380
tactaaacaa attgtgtgag aactattgag aaaaggaaat tagtgtgggc cgactagttg 13440
aaaaatgctt accagagaat gtggattgac ctgagaatag gaggttttgg gtctgggcta 13500
aaaactgtgt ggattagcac agcttgtgaa gggaaccttc agagctctac ttgacaggaa 13560
cagagagaga gtagcagagg ctagattgat gaactaggtc taagaatggc caaactcatt 13620
cattcactta acatattttc atcacctaat atgtattggt actaagcact gaggatacaa 13680
ttgtgctaag cagtgaggat acaaagatta gtttctcagg gaatttacta tctaatgagg 13740
gaggcaaata attacagtac attgttggaa gtgcagagat gtgtcagagg tctaagaagt 13800
agttaggtgc tagaaccttc ccactcatca tgcttccctt tgcagtttcc taacaaggga 13860
gccgtacata taagagtcag ttgccaacta ggaaacagag ccaggtgaac aagcatccaa 13920
tgaggcccaa atgtagggct gcctggtaag ctgaatcagg gtgcctggct gccgcccaag 13980
cctccagagc accccgtttc caaaccaaat ctcagcagcc tagtgaagag tgacaggaaa 14040
agtaaggttt ctggggaaga ttcccaaatt cacttacaca aatagccaca tagatcagaa 14100
gggccagcct gtcacttcta gcagcagatg gaaaaagaga aacgccactt gtgactgctc 14160
attacctttc agttgtgtcc ctgcctgtta agacaatgtg gcctaatgaa tcagtcagaa 14220
tagcaactgt ctcccatgca gagggagaca ggaaaagaga aagaacagag gccagaggtc 14280
tcatgggtta aggctaaaat gtggacatcc tatagaccat tcaaattcaa tatgttggct 14340
gggcacggtg gctcacagct gtaatcccag cactttagga ggccaaggca ggcggatcat 14400
ctgagatcag gagtttgaga ccagcctggc caacatggtg aaaccccatc tccactaaaa 14460
atacaaaaat tacctgggta tggtggcaca tgcctgtagt cccaggtact tgggaggctg 14520
agacaggaga atcactt 14537

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/34758

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07H 21/02; C12Q 1/68; C12P 21/00, 19/34; A01N 43/04; G06F 17/00

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.3, 24.32, 24.33; 436/6, 91.1, 91.2;800/4; 514/44; 345/418

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Medline, West 2.0, Embase, Scisearch, Biotechds, Caplus, Canadian Patents

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,602,024 A (GERALD et al) 11 February 1997. See entire patent, especially figure 5.	13-15
X	US 5,985,616 A (PARKER et al.) 16 November 1999. See entire patent, especially SEQ ID NO: 2.	13-15
A, P	WO 00/18960 A2 (MASSACUSETTS INSTITUTE OF TECHNOLOGY) 06 April 2000. See entire reference.	4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 29 MARCH 2001	Date of mailing of the international search report 24 APR 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer CYNTHIA B. WILDER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/34758

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-4, 9-12, 18 and PS1
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Please See Extra Sheet.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4, 9-15, 18 and PS1

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/34758

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

536/23.1, 24.3, 24.32, 24.33; 436/6, 91.1, 91.2;800/4; 514/44; 345/418

BOX I. OBSERVATIONS WHERE CLAIMS WERE FOUND UNSEARCHABLE
2. Where no meaningful search could be carried out, specifically:

Claims 1-4, 9-12, 18 and PS1 could not be searched because the claims are drawn to a composition and method of genotyping comprising the polymorphic site 1 wherein a G is substituted for a T at position 154224 of the NPY gene in figure 1 or Genbank nucleotide position 31293 of the NPY gene. The NPY gene reference sequence as depicted as SEQ ID NO: 1 only has 14537 nucleotide bases and there are no positions corresponding to 154224 or 31293 as recited in Table 3 of the description (page 30). Additionally figure 1 is not represented by a sequence identifier (SEQ ID NO:). Therefore a meaningful search of the positions as claimed in claims 1-4, 9-12, 18 and PS1 cannot be obtained. Because claims 13-15 read on a fragment of the reference NPY gene sequence, a search was performed on those claims.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Groups 1-8, claim(s)1-4, 9-15 and 18, in part, drawn to method of genotyping the neuropeptide Y gene of an individual with one or more polymorphic sites (PS1-PS5, PS9, PS13, PS14), an isolated polynucleotide comprising a nucleotide sequence with a polymorphic variant for the NPY gene or a fragment thereof, and a composition comprising using one of the 8 polymorphic sites, respectively. For example, if group 1 is elected, the claims 1-4, 9-15 and 18 will be examined to the extend that they apply to PS1. If group 8 is elected, the claims 1-4, 9-15 and 18 will be examined to the extend that they apply to PS14.

Groups 9-40, claim(s) 5, in part, drawn a method for predicting the haplotype pair for the NPY gene by identifying an NPY genotype for the individual at two or more polymorphic sites PS1-PS5, PS9, PS13, PS14. For example if group I is elected , the claim 5 will be examined to the extend that it applies to PS1 and PS2. If the group 40 is elected, the claim 5 will be examined to the extend that it applies to PS13 and PS14.

Groups 41-48, claim(s) 6-8, in part, drawn to a method of identifying an association between a trait and a genotype or haplotype between one of the 8 PS sites of the NPY gene. For example, if group 41 is elected, the claims 6-8 will be examined to the extend that they apply to PS1.

Groups 49-56, claim(s)16-17, in part, drawn to a nonhuman recombinant organism comprising one of the 8 PS sites, respectively. If group 49 is elected, the claims 16-17 will be examined to the extend that they apply to PS1.

Groups 57-76, claim(s) 19-20, in part, drawn to a computer system and genome anthology (database) comprising the polymorphism data of one of the 20 haplotypes in Table 5, respectively. If group 57 is elected, the claims 19-20 will be examined to the extend that they apply to the first of the 20 haplotypes.

The inventions listed as Groups 1-76 do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The broadest recitation of the first claim product, namely a fragment of the nucleic acid of the neuropeptide Y gene (NPY)of claim 15, is known and is well characterized in the prior art. Thus the "special technical feature" is not special and is not a contribution over the prior art. Furthermore, Gerald et al. (5,602,024, 11 Feb 1997) teach fragments of the NPY gene. Parker et al. (5,985,616, 16 NOV 1999) teach fragments of the NPY gene. These fragments of the NPY gene thus meets the requirements of Groups 1-8. Hence there is not a special technical feature linking the groups.

The special technical feature of Groups 1, PS1 is lacking in groups 2-8, and 37-76. The nucleic acid of Groups 1 is not required in Group 2 such that the PS site is not required for the examination of PS 1 or visa versa. Additionally, the

different PS sites are structurally and functionally distinct as shown in Table 3 and requires different searches. Similar reasons can be set forth for the unique PS1 sites and unique pairs of PS1 sites. The non-human organisms (transgenic animals) are distinct products from the nucleic acid. Finally the computer system which requires data, does not rely upon the chemical structure of the nucleic acid or the non-human organism.

The groups (1-8), (9-40) and (41-48) are not so linked as a single inventive entity. The different groups are drawn to different methods requiring different starting reagents, different method steps and different objectives. Additionally, the method of genotying the NPY gene is not required or is necessary for predicting a haplotype pair or determining traits and visa versa.